

THE SAGE ENCYCLOPEDIA OF SOCIAL SCIENCE RESEARCH METHODS

社会科学研究方法百科全书

【第二卷】

[美]米歇尔·刘易斯-伯克（Michael S. Lewis-Beck）
[美]艾伦·布里曼（Alan Bryman）
[美]廖福挺（Tim Futing Liao）—主编

沈崇麟 赵 锋 高 勇 —主译

高 勇 李 洋 叶鹏飞 —本卷译者



SHEHUI KEXUE YANJIU FANGFA BAIKE QUANSHU



更多服务


ISBN 978-7-5689-0785-9

总定价：698.00元（共三卷）



社会科学研究方法

百科全书

【第二卷】

[美]米歇尔・刘易斯-伯克（Michael S. Lewis-Beck）
[美]艾伦・布里曼（Alan Bryman）
[美]廖福挺（Tim Futing Liao）　　主编

沈崇麟 赵　锋 高　勇　　主译

高　勇 李　洋 叶鹏飞　　本卷译者

译者（按姓氏拼音排序）

高　勇（中国社会科学院）　李　洋（北京市社会科学院）　马　妍（中国社会科学院）
沈崇麟（中国社会科学院）　王　玥（首都师范大学）　杨　可（中国社会科学院）
叶鹏飞（中国劳动关系学院）赵　锋（中国社会科学院）　郑晓娟（北京大学）

词条审定专家（按姓氏拼音排序）

郭金华（北京大学）　郭志刚（北京大学）　何江穗（中国政法大学）
柯惠新（中国传媒大学）　李　原（中国社会科学院）　渠敬东（北京大学）
沈崇麟（中国社会科学院）席仲恩（重庆邮电大学）　夏传玲（中国社会科学院）

重庆大学出版社

English language edition published by SAGE Publications Inc., A SAGE Publications Company of Thousand Oaks, London, New Delhi, Singapore and Washington D.C., © [2004] by SAGE Publications, Inc.

版贸核渝字(2012)第094号

图书在版编目(CIP)数据

社会科学研究方法百科全书. 第二卷 / (美) 米歇尔·刘易斯-伯克, (美) 艾伦·布里曼, (美) 廖福挺主编; 沈崇麟, 赵锋, 高勇主译. -- 重庆 : 重庆大学出版社, 2017.8

(万卷方法)

书名原文: The SAGE Encyclopedia of Social Science Research Methods

ISBN 978-7-5689-0785-9

Ⅰ. ①社… Ⅱ. ①米… ②艾… ③廖… ④沈… ⑤赵… ⑥高… Ⅲ. ①社会科学—研究方法 Ⅳ. ①C3

中国版本图书馆CIP数据核字(2017)第199790号

社会科学研究方法百科全书

(第二卷)

[美]米歇尔·刘易斯-伯克 [美]艾伦·布里曼 [美]廖福挺 主编

沈崇麟 赵 锋 高 勇 主译

策划编辑:雷少波 柏子康 林佳木

责任编辑:雷少波 柏子康 屈腾龙 张慧梓 姜 凤 版式设计:雷少波

责任校对:谢 芳 责任印制:张 策

*

重庆大学出版社出版发行

出版人:易树平

社址:重庆市沙坪坝区大学城西路21号

邮编:401331

电话:(023)88617190 88617185(中小学)

传真:(023)88617186 88617166

网址:http://www.cqup.com.cn

邮箱:fxk@cqup.com.cn (营销中心)

全国新华书店经销

重庆升光电力印务有限公司印刷

*

开本:787mm×1092mm 1/16 印张:38 字数:866千

2017年8月第1版 2017年8月第1次印刷

ISBN 978-7-5689-0785-9 总定价:698.00元(共三卷)

本书如有印刷、装订等质量问题,本社负责调换

本卷目录

G

H

I

J

K

L

M

N

O

P

总目录

（按汉语拼音排序）

C

D

E

F

G

H

J

K

L

M

N

O

P

Q

R

T

V

W

X

Y

Z

读者指南

因果建模

话语/谈话分析

计量经济学

认识论

民族志

多层次分析

情境效应
相依观察值
固定效应模型
混合效应模型
多层次分析
非参数随机效应模型
随机系数模型
随机效应模型

多元回归

调整的 *R* 方
最佳线性无偏估计量
Beta(β)
广义最小二乘
异方差性
交互效应
误设
多重共线性
多元回归分析
非加性
R 方
回归
回归诊断
设定
估计标准误

定性资料分析

分析式归纳
CAQDAS
持续比较法
扎根理论
原生编码
备忘录，撰写备忘录
不符案例
定性内容分析

定性研究的抽样

目的抽样
定性研究的抽样
滚雪球抽样
理论抽样

调查抽样

多阶抽样
配额样本
随机抽样
代表性样本
抽样
抽样误差
分层样本
系统抽样

度量

态度测量
双极量表
维度
双重度量
哥特曼度量
指数
李克特量表
多维度量(MDS)
最优度量
量表
度量
语义差异量表
瑟斯通度量

显著性检验

Alpha(α)，检验的显著性水平
置信区间
显著性水平
单尾检验
检验的统计效力

简单回归

调查设计

时间序列

G

博弈论(Game Theory)

博弈论将传统决策理论扩展至行动者的结果不仅取决于自身决策而且取决于他人行为的情境之中。在此意义上,它是一种特别钟爱目标指向("理性")行为(goal-directed behavior)的社会互动理论。

博弈论自1944年由约翰·冯诺伊曼和奥斯卡·摩根斯顿创立以来,已经广泛应用于经济学、政治学、心理学、社会学、生物学等学科中。博弈论的关键要素包括游戏、参与者、策略、收益、最大化。"游戏"描述了至少两个行动者(参与者)之间的一种社会互动情境,包含有游戏规则、可能性、约束条件。这种情境为参与者提供了多种决策和行动的选择("策略")。参与者的策略导致有着不同效用取值的各种后果("收益")的出现。基于以上关键要素,博弈论分析可以显示,当参与者最大化自身回报时,他将会选择何种策略。因此,博弈论隶属于数学学科,而不是心理学理论。

虽然目标指向行为是博弈论的前提,但理论结果和经验证据有时令人意外。一个著名的例子就是囚徒困境【Prisoner's Dilemma, PD】。PD是对于策略性互依情境的简单模型,与合作及冲突有关。表1显示了罗伯特·阿克塞尔罗德(Axelrod, 1984)在其电脑竞赛中使用的囚徒困境。个体而言,背叛总是比合作能得到更好的收益(因此是"优势策略")。困境在于,虽然彼此背叛比彼此合作结果更差,但是参与者都不愿意单方面改变自己的策略。基于上述特征,囚徒困境可以是对许多社会情境的有效模型,如雇佣关系、军备竞赛、自由贸易。它也表明,在那些兼具合作及冲突激励的混合动机情境下,博弈论理论分析最为有趣。

表1 囚徒困境

		参与者2	
		合 作	背 叛
参与者1	合 作	3,3	0,5
	背 叛	5,0	1,1

多数经验性的博弈论研究使用实验【Experiment】、计算机模拟【Computer Simulation】以及广义上的内容分析【Content Analysis】。在内容分析中,给定的文件如统计数据或历史资料被加以分析,以建立博弈论理论模型(例如,对于第一次世界大战开始时政治策略举措的解释)。计算机模拟可以让多种不同的策略进行竞争,来模拟重复博弈(迭代)、参数改变、演化方面的后果。实验是检验博弈理论发现的最佳方式。阿纳托尔·拉波波特和艾伯特·查马(Rapoport & Chammah, 1965)关于囚徒困境的研究是将理论、实验、模拟方面结合起来的典范。关于博弈论的概要,请

参看 Gintis（2000）。

——Bernhard Prosch

（高勇译校）

参考文献

Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic Books.

Gintis, H. (2000). *Game theory evolving*. Princeton, NJ: Princeton University Press.

Rapoport, A., & Chammah, A. (1965). *Prisoner's dilemma*. Ann Arbor: Michigan University Press.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behaviour*. Princeton, NJ: Princeton University Press.

Gamma(γ)

γ 相关系数是一种用来描述两个定序变量 X 和 Y 之间关系的非参数统计量【Nonparametric Statistics】。它由戈德曼和克鲁斯卡尔提出，它与 Kendall τ 系数（以及 Kendall τ_b 和斯图亚特提出的 τ_c）很类似（参看对称量度【Symmetric Measures】）。如果不存在并列排名（tied rank）的情况，上述系数的结果都是相同的。如果有一个变量有很多并列的排名，γ 相关系数就比 τ 系数更有用也更恰当，其取值一般要比 τ 系数更大。

在一个总体中，X 和 Y 是两个定序变量，如果数据中不存在并列排名，参数 γ 被定义为(X_i, Y_j)的序列一致的概率与序列不一致的概率之差（Siegel & Castellan, 1988）。换言之：

$$\gamma = \frac{p(X_i \text{ 和 } Y_i \text{ 序列一致}) - p(X_i \text{ 和 } Y_i \text{ 序列不一致})}{1 - p(X_i \text{ 和 } Y_i \text{ 并列})} = \frac{p(X_i \text{ 和 } Y_i \text{ 序列一致}) - p(X_i \text{ 和 } Y_i \text{ 序列不一致})}{p(X_i \text{ 和 } Y_i \text{ 序列一致}) + p(X_i \text{ 和 } Y_i \text{ 序列不一致})}$$

总体的 γ 系数可以从样本中用 $\hat{\gamma} = \frac{P-Q}{P+Q}$ 来估计，其中 P 是样本中的一致数目，Q 是样本中的不一致数目。一致数目和不一致数目的计算方式如下所示。在某一对观察值中，即个案 i 的观察值(X_i, Y_i)和个案 j 的观察值(X_j, Y_j)，如果$(X_i - X_j)(Y_i - Y_j) > 0$ 则两者一致，这表明 $X_i < X_j$ 且 $Y_i < Y_j$，或者 $X_i > X_j$ 且 $Y_i > Y_j$。类似地，如果$(X_i - X_j)(Y_i - Y_j) < 0$，则两者不一致，这表明 $X_i > X_j$ 但 $Y_i < Y_j$，或者 $X_i < X_j$ 但 $Y_i > Y_j$。γ 系数的取值范围为 $-1 \sim 1$：如果 $P = 0$，则 γ 系数为-1；如果 $Q = 0$，则 γ 系数为 1。

可以用如下情境来说明这个统计量的应用，两个评委（X 和 Y）根据 10 名学生的创意写作来对他们排名，1 代表最低排名，10 代表最高排名。为了计算 γ 系数，我们先把评委 X 的评分进行升序排列，如表 1 中第二列所示。表 1 中第三列是评委 Y 给出的相应排名。接下来，我们从第一个学生开始确定每个学生的一致数和不一致数。首先比较刘大和其他 9 名学生（即刘大和陈二比较，刘大和张三比较，……，刘大和郑十比较）。一致数目（即评委 Y 给出的排名在 5 以上的学生数）为 4，不一致数目（即评委 Y 给出的排名在 5 以下的学生数）为 5。然后比较第二个学生陈二和其余 8 名学生（即陈二和张三比较，陈二和李四比较，……，陈二

和郑十比较）。一致数目（即其余学生中评委 Y 给出的排名在 4 以上的数目）为 4，不一致数目（即其余学生中评委 Y 给出的排名在 4 以下的数目）为 4。

表 1 两个评委给出的 10 名学生的创意写作排名

	评委 X	评委 Y	P	Q
刘大	1	5	4	5
陈二	2	4	4	4
张三	3	3	4	3
李四	4	2	4	2
王五	5	1	3	0
赵六	5	1	3	0
孙七	5	6	3	0
周八	6	7	1	0
吴九	6	7	1	0
郑十	7	8	0	0
$\sum$			27	14

注：P=一致数目；Q=不一致数目。

如果评委 X 给出的排名存在并列时，两个个案间不进行比较。例如，第五个学生王五就不与赵六和孙七比较，而只与周八、吴九、郑十比较。一致数目（即其余学生中评委 Y 给出的排名在 1 以上的数目）为 3，不一致数目（即其余学生中评委 Y 给出的排名在 1 以下的数目）为 0。遵循相同的程序，可以得到其余学生的 P 和 Q。

如果样本量较大，可以用以下公式对 Goodman-Kruskal γ 相关系数进行虚无假设检验：

$$z = \frac{\hat{\gamma}}{\sqrt{\frac{n(1-\hat{\gamma}^2)}{P+Q}}}$$

注意，标准误差的上限为：

$$\sqrt{\frac{n(1-\hat{\gamma}^2)}{P+Q}}$$

因此，从上述公式中得到的 z 统计量是一个保守的估计量。

——Peter Y. Chen

Autumn D. Krauss

（高勇译校）

参考文献

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.

高斯-马尔科夫定理（Gauss-Markov Theorem）

参见最佳线性无偏估计量【BLUE】；普通最小二乘【OLS】。

精神科学（Geisteswissenschaften）

这是一个 19 世纪开始被使用的术语，用于指涉人文学科和社会科学。在方法论的语境中，它也意味着在这些领域中进行研究的一种特殊方法，一种强调理解【Verstehen】作用的方法。

1849 年前，“精神科学（Geisteswissens-

chaften)”这一词语及其变体只会被零星地使用,但是从席尔把约翰·斯图亚特·密尔的《逻辑体系》翻译成德语之后,它才开始被广泛使用,席尔用术语“精神科学”来代表密尔的“道德科学(moral science)”。密尔认为心理学是最重要的道德科学,其方法应当以自然科学(naturwis-senschaften)为典范。但是,多数使用“精神科学”这一术语的德国历史学家和哲学家持有完全不同的看法。一些先前的德国作家,如歌德和黑格尔,致力于发展出一种对科学理解的概念,它不仅涵盖自然科学而且涵盖人文学科和哲学,但是他们反对把科学方法的经验主义和唯物主义看法延伸到人类社会生活的研究当中。到19世纪中期,很多领域中人们已经不再相信能够发展出单一的、综合的科学概念。相反,有影响力的哲学家认为人类社会世界的研究方法必须不同于自然科学家研究物理世界的研究方法,但是两者都具有同样的经验严格性。结果,“精神科学”这一词语又被翻译为英语,用来表示“人文研究”或“人文科学”,以便区别于在“行为科学”等标签下的社会科学实证主义观点。

对于精神科学的特殊性,最具影响力的论述之一来自威廉·狄尔泰(参见Ermarth,1978)。他认为自然科学通过研究外部特质和行为来从外部研究现象,但是我们可以从内部来理解人类及其产物(如人工制品、信仰、生活方式及其他),因为我们和研究对象拥有共同的人类本性。19世纪另一种对人文科学特殊性的有影响力的论述是由新康德主义者文德尔班和里克尔特提出的。对于他们来说,自然科学和社会科学间的主要区别不在于其主题;实际上,新康德主义热衷于强调现实只有一个(参见Hammersley,1989, pp. 28-33)。主要的区别在于概括性的科学形式和个体化的科学形式。马克斯·韦伯既吸收了狄尔泰传统也吸收了新康德主义传统,发展出了其解释(Erklären)和理解(Verstehen)并重的理解社会学,认为任何社会学论述都必须在因果和意义两个层面上是完备的(Turner & Factor, 1981)。

——Martyn Hammersley

(高勇译校)

参考文献

Ermarth, M. (1978). *Wilhelm Dilthey: The critique of historical reason*. Chicago: University of Chicago Press.

Hammersley, M. (1989). *The dilemma of qualitative method: Herbert Blumer and the Chicago school of sociology*. London: Routledge.

Mill, J. S. (1974). *Collected works of John Stuart Mill: Vol. 8. A system of logic* (Books 4-6). Toronto: University of Toronto Press.

Turner, S. P., & Factor, R. A. (1981). Objective possibility and adequate causality in Weber's methodological writings. *Sociological Review*, *29*(1), 5-28.

社会性别议题(Gender Issues)

1970年代,出于对社会科学中妇女及其生活被相对忽视的不满,研究者提出了社会性别议题。女性主义者认为,男女两性的经历差异以及两性关系的变迁对社会现象的概念化方式和研究方式具有重要意义,而人们对此还没有充分认识。现有的研究领域只是局限于生活的公共领域当中,如工作场所和学校,而对于家庭等私人领域缺乏关

注。人们错误地认为女性的经历可以简单地从男性经历中推论出来,或者认为男性的生活才是正常的或者惯常的,而女性的生活是异常的或者反常的。

社会性别与研究

处理社会性别议题的一种方式是分别处理女性和男性,然后测量他们之间的差异程度,如在健康和疾病上的差异,或者在教育获得上的差异。这样一种方式尽管有用,但是批评者认为其并不全面,因为它没有认可男女两性的经验在质量和本质上的差异。另一种方法是将女性研究添加补充到男性研究上,这是性别觉醒的早期阶段的研究特点。这种方法在就业和媒体等领域产生了非常有价值的著作,但是也同样遭到了批评,因为它并没有充分地挑战渗透于这些研究中的基本假定和基本概念。批评者认为,这些研究仍然是基于男性感知和男性利益基础上的,女性只不过是事后被补充追加了进去。这些批评导向了三种发展方向。首先,有必要研究那些对女性特别相关的其他议题。新引入的主题包括家务劳动、母亲身份、分娩生产、性感受、色情作品、男性暴力等。其次,社会科学中一些核心的视角和概念需要进行修正或改变。这引起了重要的争辩,如社会阶级对于男女两性的相对意义和重要性,如工作和休闲之间的区分是否基于性别假定等。第三种发展与社会研究的实际方法和程序有关。女性主义者中对此有着激烈的争论,有人认为定性研究风格而非定量研究风格更适于揭示女性生活中先前被遮蔽的部分。关于独立的女性主义方法是否可取或者是否可能,也有很多的争论(Ramazanoglu, 2002)。

对社会性别进行定义和操作化

对社会性别最通常的定义是把它和生理性别(biological sex)区分开来。生理性别代表的是解剖和生理上的差异,它定义了生理上的男女两性;社会性别指的是男性特质(masculinity)和女性特质(femininity)的社会建构(Connell, 2002)。社会性别并不是生理性别的直接产物,它是在特定历史时期被特定文化各有特色地创造并加以阐释的。但是,从这种用法中也产生出了几个问题。首先,性别关系的早期研究主要关注妇女和女童,以补偿其以往受到的忽视。直到最近,男性的社会性别本质,即男性特质,才成为关注的重点。此外,社会性别这一术语的常识用法也被提出了异议。有研究者认为实践中并不是只存在两种互斥的性别类型,这一点已经在国际运动员分类中表明(Ramazanoglu, 2002)。双性和变性瓦解了僵硬的性别边界,表明身体差异还是很重要的。基于二分法的定义也忽略了女性和男性内部的差异模式,例如男性特质既有暴力式的,也有非暴力式的。另一种关切认为需要把社会性别和其他差异形式联系起来,如种族、性取向、残疾、阶级、年龄等。这些类型和符号表明社会性别总是受到人的亲身经历的影响。

后现代的论点则认为社会性别根本没有任何事先给定和本质性的存在。它的存在只是因为它不断地被实施。社会性别是形形色色的、不断变动的、反复无常的,而非稳定和同质的身份。但是这样一种立场也就清除了社会性别议题当初立足的基础。尽管存在上述问题,社会性别研究者必须有一个更为具体的出发点。这导致一些评论者认为要开展有关"性别化生活(gendered lives)"的研究(Ramazanoglu, 2002, p.5),或者需要关注性别"关系"(Connell, 2002, p.9),而不是把社会性别当作一个给定的东西。这使得理解社会性别的构成内容成了具体经验研究的一部分。这也使得社会性别的构成内容成为个体和群体行动于其中

的社会关系模式的一部分。核心命题是对社会性别不能进行普遍化的判断，或者不能在研究之前就下判断。

对研究过程的影响

也有人认为社会性别可以影响研究过程，不论研究题目是什么，也不论研究视角是女性主义的还是非女性主义的。例如，两性对于社会世界的不同感知会影响为什么选择一个研究项目，这个研究项目以何种方式来组织，以及接下来提出的研究问题。研究者的性别，以及年龄和种族等其他特征，会影响到和研究参与者的互动以及收集到的信息的质量。性别在资料分析和阐释中也很重要，因为男性和女性基于其各自的经历和世界观来“阅读”资料，其方式是有差异的。这些性别影响可以在多大程度上最小化，在文献中对此有争论。有人认为性别影响对于研究来说无关紧要，也有人认为它们是偏差的可能来源。对于后者来说，有必要反思社会性别介入研究实践的方式，并且尽可能对此进行评述，以便让这些成为任何评估过程的一部分。

——Mary Maynard

（高勇译校）

参考文献

Connell, R. W. (2002). *Gender*. Cambridge, UK: Polity.

Ramazanoglu, C. (with Holland, J.). (2002). *Feminist methodology*. London: Sage.

一般线性模型（General Linear Models）

一般线性模型提供了一个基本框架，这个框架是社会科学家碰到的许多统计分析的基础。它是许多单变量、多变量以及重复测量处理方法的基础，包括 t 检验【t-Test】、（多变量）方差分析【(Multivariate) Analysis of Variance】、（多变量）协方差分析【(Multivariate) Analysis of Covariance】、多元回归分析【Multiple Regression Analysis】、因子分析【Factor Analysis】、典型相关分析【Canonical Correlation Analysis】、聚类分析【Cluster Analysis】、判别分析【Discriminant Analysis】、多维度量【Multidimensional Scaling】、重复测量设计【Repeated-measures Design】等。它可以用回归类型的公式来简洁地表达一般线性模型，例如，随机因变量 Y（如收入）的期望值可以表达为许多自变量 X_k（$k=1,\cdots,K$）（如年龄、教育、社会阶层）的函数：

$$\begin{aligned} E(Y) &= f(X_1,\cdots,X_K) \\ &= \beta_0+\beta_1 X_1+\cdots+\beta_K X_K \qquad (1) \end{aligned}$$

一般线性模型里的所有方法都可以用形式（1）给出。例如，如果所有解释变量都是类别因子，因此可以用虚拟变量【Dummy Variable】编码为 X_k 时，形式（1）给出的就是一个典型的方差分析模型。一般线性模型中的参数可以用普通最小二乘【Ordinary Least Squares】来获得。根据高斯-马尔科夫定理，如果一般线性模型的所有假定【Assumption】都得到满足，此种估计是最优的，因为此时最小二乘估计量是总体参数 β_k 的最佳线性无偏估计量【Best Linear Unbiased Estimator】。

一般线性模型和广义线性模型【Generalized Linear Model】的基本区别在于模型中的因变量类型不同。前者要求连续

变量,后者没有这种要求。另一个区别在于因变量的分布,一般线性模型期望其服从正态分布,广义线性模型中可以服从非正态分布(如二项分布【Binomial】、多项分布【Multinomial】或泊松分布【Poisson Distribution】)。另一个区别在于响应变量与解释变量之间的关系方式不同。在广义线性模型中,因变量和自变量之间的关系通过连接函数【Link Function】加以设定,连接函数的特例就是(1)中的恒等连接,这是一般线性模型中使用的唯一连接函数。但是广义线性模型中的连接函数可以采用非线性连接,如 Probit, Logit, Log 等;连接函数可以正式地表达为(2)中的反函数,或者$f^{-1}[E(Y)]$:

$$E(Y)=f(X_1,\cdots,X_K)=f(\beta_0+\beta_1X_1+\cdots+\beta_KX_K) \quad (2)$$

在此,恒等函数只是多种可能性之一,一般线性模型(1)可以视为广义线性模型(2)的特例。

一般线性模型的另一种扩展是广义加法模型【Generalized Additive Models】,它允许非参数函数从而扩展了广义线性模型。广义加法模型可以类似地用 Y 的期望值表示:

$$E(Y)=f(X_1,\cdots,X_K)=f[s_0+s_1(X_1)+\cdots+s_K(X_K)] \quad (3)$$

在此,$s_K(X_K)$是用非参数法估计出来的平滑函数。在(3)中,不仅连接函数$f^{-1}[E(Y)]$可以采用各种非线性形式,而且还用诸如 B-spline 的平滑因子来拟合每个 X_K 变量。

对于(1)和(2)中的模型还可以有另一种扩展,这种扩展反映了层级线性模型【Hierarchical Linear Model】或者多水平分析【Multilevel Analysis】模型的本质。前面我们都省略了代表个案的下标 i。如果下标 i 用来表示第一水平单元,j 表示第二水平即更高水平的单元,那么变量 Y_{ij}和 X_{ij}代表了来自两个水平的观察;此外,系数 β_{jk}现在是随机的,在不同的第二水平单元中有所不同。

一些常见的统计软件包,如 SAS 和 SPSS,都可以用灵活的程序进行一般线性模型的数据分析。SAS 和 STATA 都有程序来拟合广义线性模型和广义加法模型,也可以拟合一些层级(非)线性模型。

——Tim Futing Liao

(高勇译校)

概推性(Generalizability)

参见外在效度【External Validity】。

概化理论(Generalizability Theory)

概化理论提供了一种广泛的概念框架和一套统计工具,以量化和解释测量对象的得分中存在的一致和不一致。某种程度上,这种理论可以视为通过应用方差分析【Analysis of Variance, ANOVA】程序来扩展经典测试理论。经典理论认为观察到的得分可以分解为真实分【True Score】和一个不加分化的随机误差【Random Error】项。与此不同,概化理

论用全域分(universe score)代替了真实分,它比经典理论更为灵活,因为 ANOVA 法能够让研究者厘清误差的不同来源,而经典理论对这些不同来源的误差项不加区分。

概化理论最重要和最独有的特征可能是它的概念框架,它关注特定类型的研究和全域。一个"概化(generalizability, G)"研究就是从一个由调查者定义的各个"侧面(facets)"构成的"可采信观察的全域(universe of admissible observations)"中收集一个数据样本。通常,一个概化研究的主要结果就是对可采信的观察全域估计出一组随机效应的方差分解。一个"决策(decision, D)"研究则要估计出全域分方差(universe score variance)、误差方差、诸如信度系数【Reliability Coefficient】这样的指标,以及与调查者界定的"概化的全域(universe of generalization)"相关联的测量程序的其他统计量。

在单变量概化理论中,每一个测量对象都只有一个全域分。在多变量概化理论中,每个测量对象有多个全域分。单变量概化理论通常使用随机效应模型,而混合效应模型【Mixed-Effects Model】与多变量概化理论有更密切关联。

历史发展

概化理化的开创性研究始于李·克龙巴赫、戈尔迪纳·格勒塞尔、哈林德尔·纳德和纳基斯瓦瑞·拉贾拉特纳姆等人(Crobach, Gleser, Nanda, & Rajaratnam, 1972)1972 年发表的论著《行为测量的可靠性》(*The Dependability of Behavioral Measurements*)。罗伯特·布伦南(Brennan, 2001)在 2001 年发表的《概化理论》(*Generalizability Theory*)对这一理论进行了最新扩展。单变量概化理论的主要内容是在 1960—1961 年的技术报告中完成的,这些报告后来被修订为三篇期刊论文,分别以不同的第一作者发表(Cronbach, Gleser, Nanda, & Rajaratnam)。多变量概化理论在随后的十年中发展出来。罗纳德·费希尔、西里尔·伯特、西里尔·霍伊特、罗伯特·埃贝尔、林德奎斯特等在 1920—1955 年的研究影响了概化理论的发展。

示例

假如有一位名叫史密斯的调查者想构建一个测量程序来评估人的写作技能。首先,史密斯需要确定文章题目以及潜在的评审人。这样做也就确定了可采信的观察全域中的各个侧面。假定这些侧面都是无穷的,也就是说在理论上,任何人 p(即测量对象)可以就任何题目 t 写文章,这些文章可以被任何评审人 r 评审。如果这样,可能的数据收集设计就是 $p\times t\times r$,这里的"×"应被称为"相交"。相关的线性模型为:

$$X_{ptr} = \mu + v_p + v_t + v_r + v_{pt} + v_{pr} + v_{tr} + v_{ptr}$$

其中的 v 表示独立的得分效应(uncorrelated score effects)。这个模型有如下结果:

$$\sigma^2(X_{ptr}) = \sigma^2(p) + \sigma^2(t) + \sigma^2(r) + \sigma^2(pt) + \sigma^2(pr) + \sigma^2(tr) + \sigma^2(ptr) \quad (1)$$

它将观察得分的总方差分解成为七个方差成分(variance components),通常用随机效应 ANOVA 中的期望均方【Mean Squares】来估计这些方差成分。

假定基于 n_t 个题目和 n_r 个评审人,从史密斯的概化研究中估计出了如下方差成分:

$$\hat{\sigma}^2(p) = 0.25, \quad \hat{\sigma}^2(t) = 0.06$$
$$\hat{\sigma}^2(r) = 0.02, \quad \hat{\sigma}^2(pt) = 0.15$$
$$\hat{\sigma}^2(pr) = 0.04, \quad \hat{\sigma}^2(tr) = 0.00$$
和 $\hat{\sigma}^2(ptr) = 0.12$

再假定史密斯想要将基于 n'_t篇文章和 n'_r个评审人得到的人们的平均得分推广到包括无穷篇文章和无穷个评审人的概化全域当中。这里描述的其实是一个随机效应设计为 $p \times T \times R$ 的设计研究(D 研究)。它与史密斯的概化研究的 $p \times t \times r$ 研究很相似,但是 D 研究的样本量不必和 G 研究的样本量相同,$p \times T \times R$ 设计关注的是每个人的平均得分。在概化全域各个侧面中一个人的期望得分被称为这个人的全域得分(universe score)。

如果史密斯的测量程序中 $n'_t = 3$ 且 $n'_r = 2$,那么 D 研究估计出的随机效应方差成分为:

$$\hat{\sigma}^2(p) = 0.25$$

$$\hat{\sigma}^2(T) = \frac{\hat{\sigma}^2(t)}{n'_t} = 0.02$$

$$\hat{\sigma}^2(R) = \frac{\hat{\sigma}^2(r)}{n'_r} = 0.01$$

$$\hat{\sigma}^2(pT) = \frac{\hat{\sigma}^2(pt)}{n'_t} = 0.05$$

$$\hat{\sigma}^2(pR) = \frac{\hat{\sigma}^2(pr)}{n'_r} = 0.02$$

$$\hat{\sigma}^2(TR) = \frac{\hat{\sigma}^2(tr)}{n'_t n'_r} = 0.00$$

和 $$\hat{\sigma}^2(pTR) = \frac{\hat{\sigma}^2(ptr)}{n'_t n'_r} = 0.02$$

其中的 $\hat{\sigma}^2(p) = 0.25$ 就是估计的全域分方差。其他方差成分代表了不同类型的误差项方差。

绝对误差 Δp 就是一个人的观察均值得分和这个人的全域得分之间的差值。对史密斯的设计和全域来说,这些误差的方差为:

$$\sigma^2(\Delta) = \sigma^2(T) + \sigma^2(R) + \sigma^2(pT) + \sigma^2(pR) + \sigma^2(TR) + \sigma^2(pTR)$$

结果是 $\hat{\sigma}^2(\Delta) = 0.02 + 0.01 + 0.05 + 0.02 + 0.00 + 0.02 = 0.12$。其平方根为 $\hat{\sigma}(\Delta) = 0.35$。它可以解释为对一个随机选择的人来说测量结果的"绝对"标准误差的估计。

相对误差 δ_p 的定义为一个人的观察偏差得分和这个人的全域偏差得分之间的差值。对史密斯的设计和全域来说:

$$\sigma^2(\delta) = \sigma^2(pT) + \sigma^2(pR) + \sigma^2(pTR)$$

结果是 $\hat{\sigma}^2(\delta) = 0.05 + 0.02 + 0.02 = 0.09$。其平方根为 $\hat{\sigma}(\delta) = 0.30$,它可以解释为对一个随机选择的人来说测量结果的"相对"标准误的估计。

在概化理论中有两种类似信度的系数被广泛使用。一种系数被称为"概化系数(generalizability coefficient)",定义为:

$$E\rho^2 = \frac{\sigma^2(p)}{\sigma^2(p) + \sigma^2(\delta)}$$

它可以类比于经典理论中的信度系数。另一个系数称为"可靠性指数(index of dependability)",定义为:

$$\Phi = \frac{\sigma^2(p)}{\sigma^2(p) + \sigma^2(\Delta)}$$

由于 $\sigma^2(\delta) < \sigma^2(\Delta)$,必然有 $E\rho^2 > \Phi$。例如,使用史密斯的数据,$E\hat{\rho}^2 = 0.25/(0.25 + 0.09) = 0.74$,而 $\hat{\Phi} = 0.25/(0.25 + 0.12) = 0.68$。

——Robert L. Brennan

(高勇译校)

参考文献

Brennan, R. L. (2001). *Generalizability theory*. New York: Springer-Verlag.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: John Wiley.

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.

定性研究的概推/概推性(Generalization/Generalizability in Qualitative Research)

定性(阐述性的、民族志的)研究的发现是否应该或是否可以进行概推,这是定性研究的核心问题之一,虽然对其讨论并不充分。方法文献中对其讨论较少,原因可能是许多定性研究者认为这一问题只是那些采用定量研究方法、认为普遍规律【Nomothetic】存在的学者的一种执迷。许多定性研究者认为这并不是一个问题,因为定性研究的目标就是在具体社会情境中阐述行动者的意义,而不是测量、解释和预测。

否认概推性可能的学者认为,个体意识可以完全自主地给同一行动或情境赋予不同意义。反过来,不同的行动也可以源自相似的意义。据称"生活世界具有内在的不可决定性"(Denzin, 1983, p.133),这导致变异性太大,以至于一个具体情境中的结论不可能推广到另一个情境中。但是,这种立场的主要问题在于,事实上很难不去推广结论,因为对某一具体社会情境、事件或行为的叙述通常都会宣称具有一定典型性(或者反过来,宣称这些叙述不符合以前观察到的某些规律)。

如果概推性指的是定性研究能够产生完全概推,情境 S^{1-n} 与 S 会完全相同,或者指的是统计概推性,情境 S 发生的概率可以从 S 的情形中估计出来,那么否认概推可能性或可欲性的观点就是正确的。第一种类型的概推性除了物理学或化学中很少有可能实现,第二种类型的概推性(在社会科学中)是抽样调查研究的主题。但是,有理由认为定性研究可以进行条件性(moderatum)概推,即 S 的各个方面被视为一组更广的可识别特征下的实例(Williams, 2000)。这取决于社会环境中在文化上的一致性水平,日常生活中进行归纳推理也要以这些东西为基础,比如规则、习俗、对物理环境共同的社会构建。

从个案研究【Case Study】中就可以进行条件性概推。案例研究可以采用多种不同形式——例如,识别一种普遍社会原则的存在,分析一种社会情境,对人或事进行历时研究的扩展性案例研究等(Mitchell, 1983, pp.193-194)。案例研究的概推并不依赖于从样本到总体的统计概推(那是抽样调查中的概推方式),而是依赖于从以往理论得到的逻辑推论。理论概推因此并不试图推论总体,而是探讨理论中提出的现象的存在性。在人类学中这种推论很常见,如果在重复性的案例研究中又发现了这一现象,那么这就更加确证了理论。

如果定性研究要在政策过程中发挥作用,那么它必须要有起码的概推性。如果所有的定性研究都能给我们许多种诠释,每一种都有相同的认识论地位,那么我们如何评估社会项目或经济政策?但是,即使是从定性资料中进行条件性概推,也需要小心谨慎(正如在调查数据中要小心谨慎地处理意义问题)。即使研究发现证实了之前的理论命题,也不能肯定地认为这些发现就是典型性的。

——Malcolm Williams

(高勇译校)

参考文献

Denzin, N. (1983). Interpretive interactionism. In G. Morgan (Ed.), *Beyond method: Strategies for social research* (pp. 128-142). Beverly Hills, CA: Sage.

Mitchell, J. (1983). Case and situation analysis. *Sociological Review*, *31*(2), 186-211.

Williams, M. (2000). Interpretivism and generalisation. *Sociology*, *34*(2), 209-224.

广义加法模型(Generalized Additive Models)

加法模型是对线性回归【Linear Regression】模型 $y_i = \alpha + \sum_{j=1}^{k} X_{ij}\beta_j + \varepsilon_i$ 的改造,模型中的 y 被改造为 X 的非参数函数的加总:$y_i = \alpha + \sum_{j=1}^{k} m_j(x_{ij}) + \varepsilon_i$,其中的 ε_i 是均值为零、独立同分布的随机扰动项。广义加法模型(GAM)是对加法模型的扩展,扩展方式类似于广义线性模型(Generalized Linear Models,GLM)对于线性模型的扩展,以便处理定性因变量(如二分结果、计数数据等)。处理连续结果的加法模型可以看作 GAM 的一个特例,如同有着正态分布的普遍最小二乘【Ordinary Least Squares】回归是 GLM 的特例一样。

GAM 在探索性研究中尤其有用,或者在分析者对预测变量 X 和结果 y 之间的函数关系事先并不确定时尤其有用。在这些情形中,假定 y 的条件均值 $E(y \mid X)$是 X 的线性函数、是非常武断的;虽然线性关系很简单、易解释、简洁,但是线性函数形式可能会掩盖 $E(y \mid X)$实质上存在非常有趣的非线性关系。数据分析的动机是预测时,GAM 也很有用。通过拟合可能的非线性的 $m(\,)$ 函数,模型会抓住数据的局部特征,得到的预测会比线性模型的预测误差更小。

GAM 与其他非参数模型策略有所不同(最近的比较可以参看 Hastie, Tibshirani, & Friedman, 2001)。GAM 保留了线性回归的可加性(故而也保留了可分解性),因此更接近于 GLM,而不是神经网络【Neural Network】。GAM 关注的是条件期望的变化,因此也更接近于线性回归模型,而不是投影寻踪(projection pursuit)或树状模型方法(tree-based methods)。GAM 与处理非线性关系的参数方法也有所不同,如对变量简单取对数,用预测变量的多项式建模,或者 Box-Cox 转换;GAM 估计出来的 $m(\,)$ 函数是非参数函数,通常比参数方法更加灵活。注意,线性模型可以视为 GAM 的特例嵌套其中,这意味着可以对线性拟合的虚无假设和非参数、非线性的备择假设进行检验;即虚无假设为 $m(X) = X\beta$,换言之,对 $m(\,)$ 函数自由度为 1 的虚无假设和更高自由度的备择假设进行检验。在实践中,如果研究者有绝对把握认为某一个预测变量与 y 的关系是线性的,但是希望用更灵活的 $m(\,)$ 函数来处理其他预测变量的效应,他可以使用参数与非参数的 $m(\,)$ 函数的混合来拟合数据。

GAM 使用线性平滑(linear smoother)作为其 $m(\,)$ 函数,分别对每个预测变量依次进行。对多个预测变量拟合一个 $m(\,)$ 函数也是可能的,但是为了简化讨论,我只考虑对每个预测变量依次拟合非参数函数的情形:给定 n 个对因变量 $\mathbf{y} = (y_1, \cdots, y_n)'$ 和预测变量 $\mathbf{x} = (x_1, \cdots, x_n)'$ 的观察,一个线性平滑用目标点 x_0 处邻近的 x 和 y 观察的加权平均(即线性函数)来估计目标点 x_0 处的 $\hat{y}$。也就是说,$\hat{y} = \hat{m}(x_0) = \sum_{i=1}^{n} S(x_0, x_i; \lambda) y_i$,其中的 $S(\,)$ 是一个加权函数,参数为其中的平滑【Smoothing】参数 λ,这个参数有效地控制着目标点附近的区间宽度。区间宽度或者带宽(bandwidth)决定了用多少个加权观察 $i = 1, \cdots, n$ 来估计目标点 x_0 处的 $\hat{y}$。顾名思义,局部拟合给不在目标点 x_0 的带宽范围内数据点赋予零权重,平滑参数 λ 决定了"多近"才能算是"局部"。

实现上述想法,可以采用 LOESS 方法,这是克利夫兰(Cleveland, 1979)提出的一种常用线性平滑。给定一个目标点 x_0,loess 通过拟合一个局部加权(locally weighted)回归来得到 $\hat{y} \mid x_0 = \hat{m}(x_0)$,使用程序如下:

（1）确定 q 个 x_0 的最邻近点（nearest neighbors）（即最靠近 x_0 的 q 个 x 值）。这个集表示为 $N(x_0)$。参数 q 相当于上面定义的平滑参数 λ，一般由分析者事先设定。

（2）计算 $\Delta(x_0) = \max N(x_0) \mid x_0 - x_i \mid$，最邻近点当中离 x_0 最远的距离。

（3）为 $N(x_0)$ 中的每个点计算权数 w_i，要使用如下的三次立方（tri-cube）权重函数：

$$W\left[\frac{|x_0 - x_i|}{\Delta(x_0)}\right]$$

其中：

$$W(z) = \begin{cases}(1 - z^3)^3 & \text{当} 0 \leqslant z < 1 \\ 0 & \text{当} z < 0 \text{或} z \geqslant 1\end{cases}$$

注意，$W(x_0) = 1$，权数在选定的最邻近点中平滑地下降为零，以使 $W(x_i) = 0 \ \forall x_i \notin N(x_0)$。使用三次立方权重函数是有些武断的，任何在 $N(x_0)$ 的边境处可以平滑下降为零的权重函数都能够产生一个平滑拟合函数。

（4）使用加权最小二乘【Weighted Least Squares】和上一步中得到的权数 W 以 x 和截距项对 y 进行回归。可以使用二次型或更高阶的多项式局部回归法（Fan & Gijbels, 1996, Chap. 5），或者低阶多项式回归的混合。局部常数拟合可以得到一个简单的移动平均。

（5）从局部加权回归中得到的预测值就是平滑值 $\hat{y} \mid x_0 = \hat{m}(x_0)$。

Loess 仅是许多非参数平滑中的一种，在此用来表明其核心思想；对目标点 x_0 周围的 x 和 y 进行回归（平滑方法决定了如何对 x_0 周围的观察进行加权，即非参数回归内核的选择，以及局部回归的程度），然后报告局部回归在 x_0 处的拟合值作为非参数函数的估计值。其他常用非参数回归的平滑法包括样条【Spline】函数、Loess 及各种其他平滑函数。在各种平滑法（包括内核）中，样条法和 Loess 法是软件中最常用来执行 GAM 的方法。

对一系列目标点执行线性平滑，就会产生一个函数，x 对 y 的平滑拟合。具体而言，对每个观察进行线性平滑（即 $x_0 = x_i, i = 1, \cdots, n$）可以得到每个观察的 $\hat{y}_i$，进而得到 $\hat{\mathbf{y}} = \mathbf{S}\mathbf{y}$，其中的 $\mathbf{S}$ 是一个 $n \times n$ 的平滑矩阵，通过堆砌 n 个长度为 n 的权重向量 $S(x_i, x_i; \lambda)$ 而得到。线性平滑由此成为线性的，这类似于最小二乘回归中得到的预测值是 $\mathbf{y}$ 的线性函数：即 $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} = \mathbf{H}\mathbf{y}$，其中的 $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ 就是有名的“帽子矩阵”。平滑矩阵 $\mathbf{S}$ 与 $\mathbf{H}$ 有许多共同的性质。事实上，$\mathbf{H}$ 是线性平滑的特例，最小二乘线性回归相当于无限平滑的平滑法（正规而言，“平滑”的测量是拟合值对解释变量的二阶导数的倒数；对于线性模型来说，这个二阶导数是零）。在另一个极端，则是对数据进行内插的平滑法（这相当于对每个观察运行有一个虚拟变量的回归），此时的 $\mathbf{S} = \mathbf{I}_n$（一个 $n \times n$ 的单位矩阵），所有的非对角线元素都是零（即极端的局限拟合）。随着带宽的增加，$\mathbf{S}$ 的非对角线元素上越来越多的数目取值不是零，那些远离目标点 x_i 的数据点也会对形成 $\hat{y}_i = \hat{m}(x_i)$ 有越来越大的影响。

平滑程度与 $\mathbf{S}$ 和对角矩阵的接近程度呈现反向关系，这意味着线性平滑消耗掉的自由度可以用 $\mathbf{S}$ 的迹（trace）（对角元素的和）来近似（Hastie & Tibshirani，1990，附录 2）。由于所有的平滑法都会拟合一个截距项，$\mathrm{tr}(\mathbf{S}) - 1$ 就是对线性平滑的非参数成分消耗掉的等价自由度（equivalent degrees of freedom）的估计。这种近似提供了推断中的重要因素，平滑的残误差方差（residual error variance）σ^2 可以用下式估计：

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} [y_i - \hat{m}(x_i)]^2}{n - \text{tr}(\mathbf{S})}$$

相应地,平滑值的方差-协方差矩阵可以用 $\text{var}[\hat{\mathbf{y}}] = \mathbf{SS}'\hat{\sigma}^2$ 来估计,通过对矩阵对角线开平方可以得到逐个点的拟合平滑的标准误差。简而言之,拟合越是局部,就越不平滑,消耗的自由度就越多,$\hat{y}$的偏差就越小,但是方差就越大。

GAM 通过一种被称为后向拟合(backfitting)的迭代拟合算法【Algorithm】将线性平滑扩展到了多个预测变量的情形中,黑斯蒂和蒂布什容尼(Hastie & Tibshirani, 1990)对此进行了详尽描述。基本思路是对模型中的变量 $j=1,\cdots,k$ 进行以下循环操作,对于 y 减去截距和其他 $k-1$ 个预测变量对 x_j 的偏效应的值进行平滑,直到收敛。推广到像 GLM 那样包括定性因变量的情形中也很简单;后向拟合可以完美地嵌套于 GLM 的迭代再加权最小二乘回归算法中,这一算法由内尔德和韦德伯恩(Nelder & Wedderburn, 1972)提出。

在运用广义加法模型时的一个关键问题是,是否还有一个比拟合出来的模型更简单、更简洁的拟合。这一问题很关键,它有助于防止出现"过度拟合"(拟合了一个比数据支持程度更为花哨的函数)的情况,有助于发现各个平滑参数的合适设定。更简单的、"更少局部"的模型(包括线性拟合)可以嵌套于另一个"更为局部"的拟合中,因此似然比检验可以用于检验更为简单的模型中的约束条件是否正确这一虚无假设。这些检验在 GAM 的非参数场合中只是近似适用(首先,GAM 消耗的自由度就只是一个估计),但是在大样本中它表现良好。在统计文献中有许多自动选择平滑参数的建议方法(如 Hastie & Tibshirani, 1990, pp. 50-52),这是一个非常活跃的研究领域。

在实践中,社会科学家要注意,GAM 并不能得到线性回归或者 GLM 中得到的那种简单的单参数 x 对 y 的边缘效应。在非参数方法中,得到 x 对 y 的边缘效应的最佳方法(可能也是唯一方法)就是画出 x 对 y 的拟合曲线图。伯克和杰克曼(Beck and Jackman, 1998)的论文中提供了许多政治科学中的示例,在回归估计的传统表格之外还加入了非参数/非线性边缘效应,以及围绕拟合值的置信区间。

黑斯蒂和蒂布什容尼(Hastie & Tibshirani, 1990)的著作主题就是 GAM,关于非参数平滑(这是任何 GAM 的核心)的文献则很多。洛德(Loader, 1999)对于局部拟合方法提供了精彩的述评;沃赫拜(Wahba, 1990)对于非参数回归中样本法的使用进行了评论。S-plus 中可以估计 GAM(详尽论述请参看 Hastie, 1992)。伍德(Wood, 2001)提供了在 *R* 中的执行包,可以通过广义交叉验证来自动选择平滑参数,通过样本函数来进行平滑。

——Simon Jackman

(高勇译校)

参考文献

Beck, N., & Jackman, S. (1998). Beyond linearity by default: Generalized additive models. *American Journal of Political Science*, *42*, 596-627.

Cleveland, W. S. (1979). Robust locally-weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, *74*, 829-836.

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications*. London: Chapman & Hall.

Hastie, T. J. (1992). Generalized additive models. In J. M. Chambers & T. J. Hastie (Eds.), *Statistical models in S* (pp. 249-307). Pacific Grove, CA: Wadsworth Brooks/ Cole.

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models*. New York: Chapman & Hall.

Hastie, T. J., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning*. New York: Springer.

Loader, C. (1999). *Local regression and likelihood*. New York: Springer.

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A*, *135*, 370-384.

Wahba, G. (1990). *Spline models for observational data*. Philadelphia: SIAM.

Wood, S. (2001).mgcv: GAMs and generalized ridge regression for R. *R News*, *1*, 20-25.

广义估计方程(Generalized Estimating Equations)

广义估计方程(GEE)是广义线性模型【Generalized Linear Model】在处理重复测量(或者任何彼此相关的)数据时的一种扩展。在社会科学中,这类模型【Model】在处理定群【Panel】数据和时间序列横截面模型【Time-Series Cross-Section Model】数据时最为有用,虽然它也可以用于处理空间数据模型。

试考虑一个广义线性模型应用于重复测量数据,其中下标 $i=(1,\cdots,N)$ 表示不同单元,$t=(1,\cdots,T)$ 表明不同时点,然后有$E(Y_{it})\equiv\mu_{it}=f(\mathbf{X}_{it}\beta)$ 和 $\mathrm{Var}(Y_{it})=g(\mu_{it})/\varphi$,其中的 $\mathbf{X}$ 是一个$(NT\times k)$的协变量向量,$f(\cdot)$是连接函数【Link Function】的反函数,φ 是一个尺度参数(scale parameter)。β 的 GEE 估计值是一组 k 个微分“得分”方程的解:

$$\mathbf{U}_k(\beta)=\sum_{i=1}^{N}\mathbf{D}'_i\mathbf{V}_i^{-1}(\mathbf{Y}_i-\mu_i)=\mathbf{0} \quad (1)$$

其中,

$$\mathbf{V}_i=\frac{\mathbf{A}_i^{\frac{1}{2}}\mathbf{R}_i(\alpha)\mathbf{A}_i^{\frac{1}{2}}}{\varphi}$$

$\mathbf{A}_i$ 是一个 $t\times t$ 的对角矩阵,$g(\mu_{it})$是它的第 t 个对角元素;$\mathbf{R}_i(\alpha)$是 Y_{it}值之间的对象内条件相关(the conditional within-subject correlations)构成的 $t\times t$ 的“工作”相关矩阵(“working” correlation matrix)。工作相关矩阵的形式是由研究者决定的;但是,梁(音译)和塞格尔(Liang & Zeger,1986)表明,甚至对于 $\mathbf{R}_i(\alpha)$存在误设【Misspecification】时,β 的 GEE 估计值及其稳健方差协方差矩阵也是具有一致性的。在实践中,研究者通常会根据他们对于对象内相关性质的了解和猜测来设定 $\mathbf{R}_i(\alpha)$。

对于其他分析定群数据和时间序列横截面数据的模型来说,GEE 是一种稳健且有吸引力的替代模型。特别地,如果研究者对于对象内相关【Correlation】本身感兴趣,GEE 模型提供了这样的模型的估计方法。此类“GEE2”方法(如 Prentice & Zhao,1991)包括了一组独立的估计方程来估计成对的对象内相关。此外,GEE 能够使用一系列分布来对连续变量、二分变量、定序变量、事件发生数据进行建模。最后,许多常用的软件包都可以估计 GEE,包括 SAS,S-plus,STATA。

——Christopher Zorn

(高勇译校)

参考文献

Hardin, J. W., & Hilbe, J. M. (2003). *Generalized estimating equations*. London: Chapman & Hall.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*, 13-22.

Prentice, R. L., & Zhao, L. P. (1991). Estimating

equations for parameters in mean and covariances of multivariate discrete and continuous responses. *Biometrics*, *47*, 825-839.

Zorn, C. J. W. (2001). Generalized estimating equation models for correlated data: A review with applications. *American Journal of Political Science*, *45*, 470-490.

广义最小二乘(Generalized Least Squares)

考虑线性回归【Linear Regression】模型 $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{u}$,其中 $\mathbf{y}$ 是代表因变量观察值的 $n\times1$ 的向量,$\mathbf{X}$ 是代表自变量的 $n\times k$ 的列满秩矩阵,$\boldsymbol{\beta}$ 是代表估计参数 $k\times1$ 的向量,$\mathbf{u}$ 是代表扰动项的 $n\times1$ 的向量。根据高斯-马尔科夫定理【Gauss-Markov Theorem】,如果下列条件成立:

假定 1(A1):$E(\mathbf{u}\mid\mathbf{X})=\mathbf{0}$(即扰动项的条件均值为 0),并且

假定 2(A2):$E(\mathbf{uu}'\mid\mathbf{X})=\sigma^2\boldsymbol{\Omega}$,其中 $\boldsymbol{\Omega}=\mathbf{I}_n$,即一个 $n\times n$ 的单位矩阵(即扰动项是独立同分布的,被简称为 iid,条件方差为 σ^2)。

此时普通最小二乘【Ordinary Least Squares, OLS】估计量为 $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$,其方差-协方差矩阵 $V(\hat{\boldsymbol{\beta}}_{\mathrm{OLS}})$ 为 $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$。此时 $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ 是(a)$\boldsymbol{\beta}$ 的最佳线性无偏估计量【Best Linear Unbiased Estimator, BLUE】,因为它在所有的线性无偏估计量中有着最小的抽样变异性;(b)$\boldsymbol{\beta}$ 的一致性估计量(即当 $n\to\infty$ 时,对于任何 $\varepsilon>0$,都有 $\Pr[\mid\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}-\boldsymbol{\beta}\mid<\varepsilon]=1$;换言之 $\mathrm{plim}\,\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}=\boldsymbol{\beta}$)。

如果 A2 不成立(即 $\boldsymbol{\Omega}$ 是一个不等于单位矩阵 $\mathbf{I}_n$ 的正定矩阵),那么 $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ 仍然是无偏的,仍然是一致的,但是不再是"最佳的"。当 A2 不成立时仍依赖于 $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ 会导致错误的统计推论;A2 不成立时,$\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ 是对于 $V(\hat{\boldsymbol{\beta}}_{\mathrm{OLS}})$ 的有偏且不一致的估计量,这意味着 $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ 的估计标准误【Standard Error】是错误的,因此进行的统计推论和假设检验也是不正确的。在实践中 A2 经常是不成立的;例如,(a)将原本相异的单元汇总在一起经常会使得到的误差项具有不同的条件方差(异方差性【Heteroskedasticity】),(b)时间序列【Time-Series】数据经常会使得到的误差项不再是条件独立的(扰动项序列相关)。

当 A2 不成立时,可以进行 OLS 估计,得到最佳线性无偏估计量(至少是渐进如此)。例如,如果研究者知道与 A2 偏离的确切形式(即研究者知道 $\boldsymbol{\Omega}$),那么 GLS 估计量 $\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}=(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$ 就是最佳线性无偏估计量,其方差协方差矩阵为 $\sigma^2(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}$。注意,当 A2 成立时,$\boldsymbol{\Omega}=\mathbf{I}_n$,此时 $\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}=\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$(也就是说,OLS 是更为一般的估计量的一个特例)。

一般而言,研究者并不能确切地知道 $\boldsymbol{\Omega}$,这意味着 $\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}$ 是难以操作的,此时就要用到估计或可行广义最小二乘(estimated or feasible generalized least squares, FGLS)估计量。FGLS 估计量的实施通常包括如下几个步骤:(a)运行 OLS 分析来产生估计残差 $\hat{\mathbf{u}}$;(b)对 $\hat{\mathbf{u}}$ 进行分析来构建对 $\boldsymbol{\Omega}$ 的估计量 $\hat{\boldsymbol{\Omega}}$;(c)计算 FGLS 估计量 $\hat{\boldsymbol{\beta}}_{\mathrm{FGLS}}=(\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{y}$。第三步的执行方法通常是基于 $\hat{\boldsymbol{\Omega}}$ 可以分解为 $\hat{\boldsymbol{\Omega}}=\mathbf{P}^{-1}(\mathbf{P}')^{-1}$,然后运行 $\mathbf{y}^*=\mathbf{Py}$ 对 $\mathbf{X}^*=\mathbf{PX}$ 的加权最小二乘【Weighted Least Squares】回归;也就是说,$\hat{\boldsymbol{\beta}}_{\mathrm{FGLS}}=(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{y}^*$。

FGLS 估计量的性质取决于 $\boldsymbol{\Omega}$ 的形式

（即与 A2 中的条件独立同分布假定偏离的性质）和 $\hat{\boldsymbol{\Omega}}$ 的性质，所以不能一概而论。FGLS 估计量的有限样本性质通常可以根据具体情况通过蒙特卡洛【Monte Carlo】实验来得到；事实上，在有些对 A2 的违反并不严重的情况下，可能会发现 $\hat{\boldsymbol{\beta}}_{OLS}$ 比 $\hat{\boldsymbol{\beta}}_{FGLS}$ 更为有效（例如，Chipman，1979；Rao & Griliches，1969）。渐近【Asymptotic】结果更为丰富，通常是通过表明 $\hat{\boldsymbol{\beta}}_{FGLS}$ 和 $\hat{\boldsymbol{\beta}}_{GLS}$ 是渐近等效的，所以 $\hat{\boldsymbol{\beta}}_{FGLS}$ 是 $\boldsymbol{\beta}$ 的一致性估计量和渐近有效估计量（如 Amemiya，1985，pp. 186-222；Judge，Griffiths，Hill，& Lee，1980，pp. 117-118）。

在社会科学中，FGLS 最常见于残差自相关和异方差性的简单形式中。例如，流行的用于扰动项一阶自相关 AR(1) 问题的科克伦-奥克特（Cochrane & Orcutt，1949）和 Prais-Winsten（Prais & Winsten，1954）程序得出的就是 FGLS 估计量。FGLS 用来处理异方差性的例子在许多计量经济学教材中都可以见到：贾奇等人（Judge，et al.，1980，pp. 128-145）和艾美米亚（Amemiya，1985，pp. 198-207）都认真地探究了 $\boldsymbol{\beta}$ 的 FGLS 统计量如何用来处理常见的异方差项形式。将原本相异的单元汇总在一起会产生成组的异方差性（groupwise heteroskedasticity），由此产生的扰动项在观察组内是条件独立同分布的；这种模型事实上是泽尔纳（Zellner，1962）的相依回归【Seemingly Unrelated Regression，SUR】模型的一个特例。应用于泽尔纳的半相依回归模型的 FGLS 会得出一个跨方程的加权方案，它被用于三阶段最小二乘估计量中的第三阶段（Zellner & Theil，1962）。

如果定群【Panel】数据中存在自变量不能捕捉的因单元而异（unit-specific）或因时期而异（period-specific）的异质性【Heterogeneity】时，FGLS 也可以用于估计这类模型，例如 $y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + v_i + \eta_t + \varepsilon_{it}$，其中 v_i 和 η_t 分别是单元 i 和时期 t 所特有的误差项（或随机效应）。复合起来的误差项 $u_{it} = v_i + \eta_t + \varepsilon_{it}$ 通常都不是独立同分布的，但是它的方差-协方差矩阵 $\boldsymbol{\Omega}$ 可以用误差各组成部分的估计方差和来进行估计；给定 $E(u_{it} \mid \mathbf{x}_{it} = 0)$ 和 $E(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})$ 是满秩的条件，FGLS 提供了对 $\boldsymbol{\beta}$ 的一致性估计量（参见 Wooldridge，2002，Chap.10）。

最后 FGLS 估计量通常可以用最大似然估计【Maximum Likelihood Estimation，MLE】来获取，这样一个步骤就可以得到 $\boldsymbol{\beta}$ 的一致性估计。虽然独立同分布假定（A2）常常可以极大地简化 MLE 的推导和计算，但是通常加入少量用于决定 $\boldsymbol{\Omega}$ 的内容和结构的辅助参数（例如考虑 $\boldsymbol{\Omega}$ 的相对简单形式，如一阶残差自相关、成组的异方差性、简单的误差组成结构等），就可以处理更为常见（也更简单）的对于 A2 假定的违背。这样，用 MLE 就可以同时估计实质兴趣参数 $\boldsymbol{\beta}$ 和处理假定的误差过程的参数，由此得到的估计结果与 FGLS 估计是渐近等效的。

——Simon Jackman

（高勇译校）

参考文献

Amemiya, T. (1985). *Advanced econometrics*. Cambridge, MA: Harvard University Press.

Chipman, J. S. (1979). Efficiency of least squares estimation of linear trend when residuals are autocorrelated. *Econometrica*, *47*, 115-128.

Cochrane, D., & Orcutt, G. H. (1949). Application of least squares relationships containing autocorrelated error terms. *Journal of the American Statistical Association*, *44*, 32-61.

Judge, G. G., Griffiths, W. E., Hill, R. C., & Lee, T.-C. (1980). *The theory and practice of econometrics*. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley.

Prais, S. J., & Winsten, C. B. (1954). *Trend estimators and serial correlation*. Chicago: Cowles

Commission.

Rao, P., & Griliches, Z. (1969). Small-sample properties of several two-stage regression methods in the context of auto-correlated errors. *Journal of the American Statistical Association*, *64*, 253-272.

Wooldridge, J. M. (2002). *Econometric analysis of cross-section and panel data*. Cambridge: MIT Press.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, *57*, 348-368.

Zellner, A., & Theil, H. (1962). Three-stage least squares: Simultaneous estimation of simultaneous equations. *Econometrica*, *30*, 54-78.

广义线性模型(Generalized Linear Models)

广义线性模型(GLM)扩展了线性模型的基本结构,以便用统一的理论形式来考虑非正态和非定距的结果变量。社会科学中,常会遇到不符合线性模型标准假定的结果变量,因此发展出许多不同的回归模型:计数变量有泊松回归【Poisson Regression】;二分变量有Logit分析和Probit分析;持续期【Duration】数据有指数模型;删节【Truncation】数据有伽马模型(gamma models);等等。虽然这些模型很常用也易理解,但应用者还没有普遍认识到,几乎所有这些看似独特的回归技术都是广义线性模型的特例。

在广义线性模型中,解释变量的处理与通常方式完全一样,即建立一个线性系统部分,$\boldsymbol{\eta}=\mathbf{X}\boldsymbol{\beta}$,这个部分定义了统计模型的右端。因为结果变量的期望值,$\boldsymbol{\mu}=-\bar{\mathbf{y}}$不再来自定距测量形式,因此标准的中心极限和渐进定理不再适用,需要一个"连接"函数【Link Function】来界定数据的线性系统部分和结果变量均值间的关系,即$\boldsymbol{\mu}=g(\mathbf{X}\boldsymbol{\beta})$。如果这个连接函数是恒等函数,$g(\mathbf{X}\boldsymbol{\beta})=\mathbf{X}\boldsymbol{\beta}$,那么广义线性模型就简化为线性模型,模型也因此得名。

广义线性模型基于内尔德和韦德伯恩(Nelder & Wedderburn,1972)以及麦卡拉和内尔德(McCullagh & Nelder,1989)提出的非常成熟的理论,理论认为任何结果变量,只要其参数形式可以(在代数上)转化为指数族形式(exponential family form)的参数形式,就可以导出一个连接函数,将参数形式的均值函数和线性系统部分联系起来。指数族形式只是一种用来表示不同概率质量函数(probability mass function,PMF)(用于离散形式)和概率密度函数【Probability Density Function,PDF】(用于连续形式)的标准化途径。

GLM方法的价值在于,一系列看似不相关的非线性回归模型可以在一个统一的框架得到整合,从而寻找模型设定、数值估计、残差分析、拟合优度分析、报告等过程全都可以统一起来。这样,研究者和应用者就可以学到一种通用性的程序,适用于广泛的数据类型,还可以进一步发展,以便涵盖比我们在此描述的还要广泛的一系列假定。

指数族形式

理解广义线性模型的关键,在于了解如何用指数族形式来表达普通的连续数据的概率密度函数和离散数据的概率质量函数,指数族形式很容易生成连接函数和矩统计量。指数族形式源于费希尔(Fisher,1934, pp. 286-296),但是直到最近才被认识到,指数族的成员有着维度固定的充分统计量(fixed dimension sufficient statistics),可以得到独特的具有强一致性和渐近正态性的最大似然估计【Maximum Likelihood Estimation】(Gill, 2000, p. 22)。

随机变量 Z 有形式为 $f(z \mid \zeta)$ 的单参数的条件 PDF 或 PMF。如果它可以表示为如下形式,则可被划定属于指数族形式:

$$f(z \mid \zeta) = \exp[\underbrace{t(z)u(\zeta)}_{\text{交互部分}} + \underbrace{\log(r(z)) + \log(s(\zeta))}_{\text{加法部分}}] \quad (1)$$

其中的 r 和 t 是不依赖于 ζ 的 z 的实值函数,s 和 u 是不依赖于 z 的 ζ 的实值函数,且有 $r(z)>0, s(\zeta)>0\ \forall z,\zeta$。这里的关键是指数中子函数的独特性。被称为交互部分(interaction component)的那一部分必须是 z 的函数 $t(z)$ 与 ζ 的函数 $u(\zeta)$ 的乘积。此外,加法部分(additive component)必须也是由两个 z 的和 ζ 的子函数构成的,但是这里两者是加起来的。

指数族表达式的典范式(canonical form)是(1)式的简化,它揭示出了更大的结构,更简洁地概括了数据。这种一对一的转换是如下进行的。如有必要(即如果典范式中还不存在此转换),进行如下转换:$y = t(z)$ 和 $\theta = u(\zeta)$,现在表达式(1)变为:

$$f(y \mid \theta) = \exp[y\theta - b(\theta) + c(y)] \quad (2)$$

式(2)中的子函数有着特定的标志:y 是数据的典范式(通常也是一个充分统计量),θ 是未知参数的自然典范式,$b(\theta)$ 被称为"累积量函数"或"正态化常数"。函数 $c(y)$ 在估计过程中通常并不重要。但是,$b(\theta)$ 在决定均值和方差函数以及连接函数上都起着重要作用。还要注意,对于随机抽样来说,经典形式具有不变性,这意味着对于独立同分布数据来说,它可以保持其函数形式:

$$f(y \mid \theta) = \exp\left[\sum_{i=1}^{n} y_i\theta - nb(\theta) + \sum_{i=1}^{n} c(y_i)\right]$$

通过一个长度为 k 的参数向量 $\boldsymbol{\theta}$ 可以扩展到多个参数的情形中:

$$f(y \mid \boldsymbol{\theta}) = \exp\left\{\sum_{j=1}^{k}[y\theta_j - b(\theta_j)] + c(y)\right\}$$

作为上述理论的示例,我们可以把常见的二项分布的 PMF 改写为指数族形式。传统的二项分布的 PMF 为:

$$f(y \mid n,p) = \binom{n}{y} p^y (1-p)^{n-y}$$

改写后的指数族形式为:

$$f(y \mid n,p) = \exp\left\{\underbrace{y\log\left(\frac{p}{1-p}\right)}_{y\theta} - \underbrace{[-n\log(1-p)]}_{b(\theta)} + \underbrace{\log\binom{n}{y}}_{c(y)}\right\}$$

这里,子函数都加了标签,以对应于前面讨论过的典范式。

指数族形式与最大似然

在给定的数据值观察矩阵 $f(\boldsymbol{\theta} \mid \mathbf{X})$ 下,要获得对未知的 k 维向量 $\boldsymbol{\theta}$ 的估计值,我们可以用将似然函数最大化的标准技术来发现向量 $\boldsymbol{\theta}$ 的"最可能"取值(Fisher, 1925, pp. 707-709)。渐进理论告诉我们,对于足够大的样本而言,指数族形式的似然平面在各个维度上是单峰的,因此最大似然估计过程等同于发现 k 维的众数(mode)。

将 $f(\mathbf{X} \mid \boldsymbol{\theta})$ 视为给定观察数据 $\mathbf{X}$ 条件下 $\boldsymbol{\theta}$ 的函数,那么 $L(\boldsymbol{\theta} \mid \mathbf{X}) = f(\mathbf{X} \mid \boldsymbol{\theta})$ 被称为似然函数。最大似然估计 $\hat{\boldsymbol{\theta}}$ 有如下性质:$L(\hat{\boldsymbol{\theta}} \mid \mathbf{X}) \geq \mathbf{L}(\boldsymbol{\theta} \mid \mathbf{X})\ \forall \boldsymbol{\theta} \in \boldsymbol{\Theta}$ 其中的 $\boldsymbol{\Theta}$ 是参数形式假定中所允许的 $\boldsymbol{\theta}$ 的取值范围。

似然函数的自然对数更易于处理,这不会改变任何参数结果,因为似然函数和对数似然函数的最大值是相同的。在式(2)中

再加入一个比例参数(scale parameter),指数族典范式的似然函数为:

$$\begin{aligned}\ell(\theta,\psi \mid \mathbf{y}) &= \log(f(\mathbf{y} \mid \theta,\psi)) \\ &= \log\left\{\exp\left[\frac{\mathbf{y}\theta - b(\theta)}{a(\psi)} + c(\mathbf{y},\psi)\right]\right\} \\ &= \frac{\mathbf{y}\theta - b(\theta)}{a(\psi)} + c(\mathbf{y},\psi) \qquad (3)\end{aligned}$$

由于所有项都可以指数族形式的指数来表示,对数似然可以去掉这种指数形式,而得到一个简洁的表达式。对数似然函数对目标参数的一阶导数就是得分函数(score function)。从此时起,比例参数 ψ 被处理为冗余参数(nuisance parameter)(不是主要关注对象),我们将其估计为一个标量。由此得到的得分函数用 $\dot{\ell}(\theta \mid \psi,\mathbf{y})$ 表示,公式如下:

$$\begin{aligned}\dot{\ell}(\theta,\psi \mid \mathbf{y}) &= \frac{\partial}{\partial\theta}\ell(\theta \mid \psi,\mathbf{y}) \\ &= \frac{\partial}{\partial\theta}\left[\frac{\mathbf{y}\theta - b(\theta)}{a(\psi)} + \mathbf{c}(\mathbf{y},\psi)\right] \\ &= \frac{\mathbf{y} - \frac{\partial}{\partial\theta}b(\theta)}{a(\psi)} \qquad (4)\end{aligned}$$

最大似然法的机制,是令 $\dot{\ell}(\theta \mid \psi,\mathbf{y})$ 等于零,然后解出目标参数,给出最大似然估计 $\hat{\theta}$。此外,似然原则(likelihood principle)表明,一旦观察完了数据,数据中所有用于估计$\hat{\theta}$的可得证据都被包括在计算出的似然函数 $\ell(\theta \mid \psi,\mathbf{y})$ 中。

确定函数 $b(\theta)$ 的价值,在于它和均值函数及方差函数有着直接联系。用(3)中的符号表示出式(4)的期望值,则有如下结果:

$$E_Y\left[\frac{y - \frac{\partial}{\partial\theta}b(\theta)}{a(\psi)}\right] = 0$$

$$\int_Y yf(y)\,\mathrm{d}y - \int_Y \frac{\partial b(\theta)}{\partial\theta}f(y)\,\mathrm{d}y = 0$$

$$\underbrace{\int_Y yf(y)\,\mathrm{d}y}_{E(Y)} - \frac{\partial b(\theta)}{\partial\theta}\underbrace{\int_Y f(y)\,\mathrm{d}y}_{1} = 0$$

$$\Rightarrow E[Y] = \frac{\partial}{\partial\theta}b(\theta)$$

上述过程中的第二步到最后一步需要积分范围内必须满足一般正则条件(general regularity condition),所有的指数族分布都满足上述条件(Gill, 2000, p.23)。这意味着式(2)也给出了相关的指数分布族的均值:$\mu = b'(\theta)$。可以证明,$b(\theta)$ 的二阶导数就是方差函数(Gill, 2000, p.26)。然后(如有必要)乘上比例参数,再对典范参数进行替换,就得到了标准的方差计算式。表 1 中列出了二项分布 PMF 的上述计算式:

表 1 二项分布的均值和方差函数

$b(\theta)$	$n\log(1+\exp(\theta))$
$E[Y] = \frac{\partial}{\partial\theta}b(\theta)\big\vert_{\theta=u(\zeta)}$	np
$\frac{\partial^2}{\partial\theta^2}b(\theta)$	$n\exp(\theta)(1+\exp(\theta))^{-2}$
$\mathrm{Var}[Y] = a(\psi)\frac{\partial^2}{\partial\theta^2}b(\theta)\big\vert_{\theta=u(\zeta)}$	$np(1-p)$

广义线性模型定理

我们从满足如下高斯-马尔科夫条件的标准线性模型讲起:

$$\begin{aligned}\underset{(n\times1)}{\mathbf{V}} &= \underset{(n\times p)(p\times1)}{\mathbf{X\beta}} + \underset{(n\times1)}{\boldsymbol{\varepsilon}} \\ E(\underset{(n\times1)}{\mathbf{V}}) &= \underset{(n\times1)}{\boldsymbol{\theta}} = \underset{(n\times p)(p\times1)}{\mathbf{X\beta}} \qquad (5)\end{aligned}$$

式(5)两个方程的右端包括如下部分:$\mathbf{X}$ 为观察数据值的矩阵;$\mathbf{X\beta}$ 为“线性结构向量”;$\boldsymbol{\varepsilon}$ 为误差项。左端包括均值向量(即系统部分)$E(\mathbf{V})=\theta$。变量 $\mathbf{V}$ 是独立同分布的正态分布,均值为 θ,方差为 σ^2。现在我们

基于结果变量(outcome variable)的均值用一个新的"线性预测值"对上式进行扩展,其中的结果变量不再必须服从正态分布或必须是连续变量:

$$\underset{(n\times1)}{g(\boldsymbol{\mu})} = \underset{(n\times1)}{\boldsymbol{\eta}} = \underset{(n\times p)(p\times1)}{\mathbf{X}\boldsymbol{\beta}}$$

此处的关键是,$g(\ )$是均值向量 $\boldsymbol{\mu}$ 的可逆平滑(smooth)函数。

在模型中,解释变量的效应是通过从线性结构 $\mathbf{X\beta}$ 到线性预测值 $\boldsymbol{\eta}=g(\boldsymbol{\mu})$ 的联系体现出来的,这一联系决定于连接函数 $g(\ )$的形式。这一连接函数将线性预测值与结果变量的均值联系起来,而非与结果变量本身直接相联系,这样结果变量就可以有多种非正态的分布形式。连接函数通过均值函数把服从各种分布的响应变量的随机部分与支持线性系统部分的标准正态理论联系起来:

$$g(\boldsymbol{\mu})=\boldsymbol{\eta}=\mathbf{X\beta}$$
$$g^{-1}(g(\boldsymbol{\mu}))=g^{-1}\boldsymbol{\eta}=g^{-1}(\mathbf{X\beta})=\boldsymbol{\mu}=E(\mathbf{Y})$$

此外,这一连接函数可以从指数族中的子函数 $b(\theta)$ 中以均值函数形式得到。

例如,对于二项分布的 PMF 来说,我们看到 $b(\theta)=n\ \log(1+\exp(\theta))$ 和 $\theta=\log\left(\frac{p}{1-p}\right)$。以连接函数的符号重新表达前式,可以得到 $\boldsymbol{\eta}=g(\boldsymbol{\mu})=\log\left(\frac{\boldsymbol{\mu}}{1-\boldsymbol{\mu}}\right)$,或者可以给出其反联系形式 $\boldsymbol{\mu}=g^{-1}(\boldsymbol{\eta})=\frac{\exp(\boldsymbol{\theta})}{1+\exp(\boldsymbol{\theta})}$。

当前所述的线性模型的扩展包括如下组成部分:

(1)随机部分(stochastic component):$\mathbf{Y}$ 是随机部分,它是独立同分布的,服从某种特定的指数族分布,均值为 $\boldsymbol{\mu}$。

(2)系统部分(systematic component):$\boldsymbol{\eta}=\mathbf{X\beta}$ 是系统部分,具有相关的高斯-马尔科夫正态基础。

(3)连接函数(link function):随机部分和系统部分通过函数 $\boldsymbol{\eta}$ 相联系,它可以从典范联系 $b(\theta)$ 的反函数中得出。

(4)残差(residuals):虽然残差可以与标准线性回归中相同的方式表达,即观察到的结果变量值减去预测的结果变量值,但是更有用的量是如下所述的偏差残差(deviance residual)。

广义线性模型的估计

不同于标准线性模型,广义线性模型的估计没有一个封闭解公式(closed-form analytical expression)[如$\hat{\beta}=(\mathbf{X'X})^{-1}\mathbf{Xy}$]。对未知参数的最大似然估计是通过一个称为"迭代加权最小二乘(iterative weighted least squares,IWLS)"的加权数值程序得到的,这个过程中的每个循环都会用改进的加权来提升似然值。这种程序是由内尔德和韦德伯恩(Nelder & Wedderburn,1972, pp. 372-374)首先提出的,最早在 GLIM 软件包中实现,适用于所有基于指数族形式(以及某些其他形式)的广义线性模型。目前,所有专业水准的统计计算程序都使用 IWLS 来发现广义线性模型的最大似然估计。

总策略是对正规方程使用牛顿-拉普森算法和费希尔得分算法,这与迭代性地使用加权最小二乘(并且更简单)相同。定义当前(或初始)的线性预测点为:

$$\underset{(n\times1)}{\hat{\boldsymbol{\eta}}_0} = \underset{(n\times p)(p\times1)}{\mathbf{X}'\boldsymbol{\beta}_0}$$

其中从 $g^{-1}(\hat{\boldsymbol{\eta}}_0)$ 中得到拟合值$\hat{\boldsymbol{\mu}}_0$。根据下式得到"调整后的因变量":

$$\underset{(n\times1)}{\mathbf{z}_0} = \underset{(n\times1)}{\hat{\boldsymbol{\eta}}_0} + \underset{\text{diag}(n\times n)}{\left(\frac{\partial\boldsymbol{\eta}}{\partial\boldsymbol{\mu}}\bigg|_{\hat{\boldsymbol{\mu}}_0}\right)} \underset{(n\times1)}{(\mathbf{y}-\hat{\boldsymbol{\mu}}_0)}$$

这是应用于数据的连接函数的线性化形式。作为这一导函数的示例,二项式(the binomial form)如下所示:

$$\boldsymbol{\eta} = \log\left(\frac{\boldsymbol{\mu}}{1-\boldsymbol{\mu}}\right) \Rightarrow \frac{\partial\boldsymbol{\eta}}{\partial\boldsymbol{\mu}} = (\boldsymbol{\mu})^{-1}(1-\boldsymbol{\mu})^{-1}$$

然后构建二次权重矩阵（quadratic weight matrix），它是 $\mathbf{z}$ 的方差：

$$\underset{\text{diag}(n\times n)}{\omega_0^{-1}} = \left(\underset{\text{diag}(n\times n)}{\frac{\partial\boldsymbol{\eta}}{\partial\boldsymbol{\mu}}\bigg|_{\hat{\mu}_0}}\right)^2 \underset{\text{diag}(n\times n)}{v(\boldsymbol{\mu})\,|_{\hat{\mu}_0}}$$

其中的 $v(\boldsymbol{\mu})$ 是如下方差函数：$\frac{\partial}{\partial\boldsymbol{\theta}}b'(\boldsymbol{\theta}) = b''(\boldsymbol{\theta})$。这一过程一定是迭代的，因为 $\mathbf{z}$ 和 $\boldsymbol{\omega}$ 都取决于当前的拟合值 $\boldsymbol{\mu}_0$。总过程可以总结为三个步骤：

（1）构建 $\mathbf{z}$ 和 $\boldsymbol{\omega}$。使用权重和协变量对 $\mathbf{z}$ 进行回归，得到一个新的暂时估计：

$$\underset{(p\times1)}{\hat{\boldsymbol{\beta}}_1} = (\underset{(p\times n)}{\mathbf{X}'}\ \underset{(n\times n)}{\boldsymbol{\omega}_0}\ \underset{(p\times p)}{\mathbf{X}})^{-1}\ \underset{(p\times n)}{\mathbf{X}'}\ \underset{(n\times n)}{\boldsymbol{\omega}_0}\ \underset{(n\times 1)}{\mathbf{z}_0}$$

（2）使用系数向量估计来更新线性预测值：

$$\hat{\boldsymbol{\eta}}_1 = \mathbf{X}'\hat{\boldsymbol{\beta}}_1$$

（3）进行如下迭代：

$$\begin{aligned}
\mathbf{z}_1\boldsymbol{\omega}_1 &\Rightarrow \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\eta}}_2 \\
\mathbf{z}_2\boldsymbol{\omega}_2 &\Rightarrow \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\eta}}_3 \\
\mathbf{z}_3\boldsymbol{\omega}_3 &\Rightarrow \hat{\boldsymbol{\beta}}_4, \hat{\boldsymbol{\eta}}_4 \\
&\vdots
\end{aligned}$$

在极普遍的条件下，只要分布满足指数族，迭代加权最小二乘程序都能够发现似然函数的最大值，由此得到未知参数向量的最大似然估计 $\hat{\beta}$。此外，由 $\hat{\sigma}^2(\mathbf{X}'\boldsymbol{\Omega}\mathbf{X})^{-1}$ 产生的矩阵在概率上收敛于 $\hat{\beta}$ 的方差矩阵。

残差与偏差

广义线性模型的一大优势就是它不再要求残差必须均值为零、方差恒定这样的高斯-马尔科夫假定。但是，要得到这种优势，就必须付出解释更复杂的随机结构的代价。从线性模型中计算出来的响应（response）残差向量 $\mathbf{R}_{响应} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$，必须被更替为包含 GLM 连接函数的形式，即 $\mathbf{R}_{响应} = Y - g^{-1}(\mathbf{X}\boldsymbol{\beta})$，但是它并不能提供如我们从标准线性模型中得到的那种很好的分布理论。

对于广义线性模型来说，更有用的但与此相联系的一个概念是偏差函数（deviance function）。它的构建方式类似于似然比统计量，即比较从所提议的模型设定得到的对数似然与从饱和模型中得到的最大对数似然（饱和模型使用相同数据、相同连接函数，有 n 个数据点和 n 个设定参数）。得到的差值乘以 2，被称为加总偏差（summed deviance）（Nelder & Wedderburn, 1972, pp. 374-376）。这种拟合优度测量的直觉道理在于，这个加总是单个似然贡献与原始数据对饱和模型的贡献进行比较的加总。这里的要点在于，所提议的模型的对数似然

$$\ell(\hat{\boldsymbol{\theta}}, \psi \mid \mathbf{y}) = \sum_{i=1}^{n} \frac{y_i\hat{\boldsymbol{\theta}} - b(\hat{\boldsymbol{\theta}})}{a(\psi)} + c(\mathbf{y}, \psi)$$

要与相同数据、相同连接函数，但是有 n 个数据点和 n 个参数的对数似然函数（即饱和模型对数似然函数）相比：

$$\ell(\tilde{\boldsymbol{\theta}}, \psi \mid \mathbf{y}) = \sum_{i=1}^{n} \frac{y_i\tilde{\boldsymbol{\theta}} - b(\tilde{\boldsymbol{\theta}})}{a(\psi)} + c(\mathbf{y}, \psi)$$

后者是给定数据的对数似然函数所能得到的最大值。偏差函数的定义如下：

$$\begin{aligned}
D(\boldsymbol{\theta}, \mathbf{y}) &= 2\sum_{i=1}^{n}\left[\ell(\tilde{\boldsymbol{\theta}}, \psi \mid \mathbf{y}) - \ell(\hat{\boldsymbol{\theta}}, \psi \mid \mathbf{y})\right] \\
&= 2\sum_{i=1}^{n}\left[y_i(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) - b(\tilde{\boldsymbol{\theta}}) - b,\right. \\
&\quad \left.\psi(\hat{\boldsymbol{\theta}})\right]a(\psi)^{-1}
\end{aligned}$$

这一统计量渐进服从自由度为 $n-k$ 的卡方分布[虽然如果结果变量的离散粒度(discrete granularity)很高,而且数据量不大,这一分布假定就不太成立了]。庆幸的是,对于许多指数族来说,这些偏差函数通常都已经进行了制表,不需要进行分析性计算了。对于二项分布的情形来说,偏差函数为:

$$D(m,p) = 2\sum\left[y_i \log\left(\frac{y_i}{\mu_i}\right) + (n_i - y_i)\log\left(\frac{n_i - y_i}{n_i - \mu_i}\right)\right]$$

其中的饱和对数似然在拟合 $p_i = y_i / n_i$ 时取得其最大值。

我们可以用与线性模型残差相类似的方式得到个体的偏差贡献。单个点的偏差函数就是第 y_i 个点的偏差函数:

$$d(\boldsymbol{\theta}, y_i) = -2\{y_i(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) - [b(\tilde{\boldsymbol{\theta}}) - b(\hat{\boldsymbol{\theta}})]\}a(\psi)^{-1}$$

在这一点上的偏差残差按如下方式计算:

$$\mathbf{R}_{偏差} = \frac{y_i - \mu_i}{|y_i - \mu_i|}\sqrt{|d(\boldsymbol{\theta}, y_i)|}$$

其中的 $\frac{y_i - \mu_i}{|y_i - \mu_i|}$ 只是一个保留其正向号的函数。

示例:多项响应模型

如下例子是二项分布的一个扩展,其中定义结果变量的不只有两种可能事件,而是有 k 个事件。个体 i 的结果由长度为 $k-1$ 的向量 $y_i = [y_{i1}, y_{i2}, \cdots, y_{i(k-1)}]$ 表示,这一向量除了那个表示其响应的元素外全部为零,或者对于选择了参考类别项的个体来说,这一向量的元素全部为零。如果 $k=2$,则它就简化为二项结果。

在此,目标是从包括 $n\times(k-1)$ 的结果矩阵 y,$n\times p$ 的协变量矩阵 $\mathbf{X}$(起始列全部元素为 1)的数据集中,对样本量 n 估计出长度为 $k-1$ 的类别概率 $\boldsymbol{\mu} = g^{-1}(\boldsymbol{\eta}) = \boldsymbol{\pi} = [\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \cdots, \boldsymbol{\pi}_{k-1}]$。这一设定的 PMF 是多项分布,要用 logit(有时用 probit)连接函数进行估计,要对 $k-1$ 个类别分别给出第 i 个人选择类别 r 的概率:

$$P(\mathbf{y}_i = r \mid \mathbf{X}) = \frac{\exp(\mathbf{X}_i \boldsymbol{\beta}_r)}{1 + \sum_{s=1}^{k-1}\exp(\mathbf{X}\boldsymbol{\beta}_s)}$$

其中的 β_r 是第 r 个类别的系数向量。

表 2　分上学年数与性别的幸福感

(1977 年综合社会调查)

自陈状态		上学年数			
		<12	12	13~16	17+
男性	不幸福	40	21	14	3
	较幸福	131	116	112	27
	很幸福	82	61	55	27
女性	不幸福	62	26	12	3
	较幸福	155	156	95	15
	很幸福	87	127	76	15

使用的数据来源于 1977 年的综合社会调查,许多作者都分析过这个三类别的例子。表 2 是数据的汇总。由于应用中只有三个类别,第 i 个人的多项 PMF 为:

$$f(\mathbf{y}_i \mid n_i, \boldsymbol{\pi}_i) = \frac{n_i}{y_{i1}!\ y_{i2}!\ (n - y_{i1} - y_{i2})!}\boldsymbol{\pi}_{i1}^{y_{i1}}\boldsymbol{\pi}_{i2}^{y_{i2}}(1 - \boldsymbol{\pi}_{i1} - \boldsymbol{\pi}_{i2})^{n - y_{i1} - y_{i2}}$$

似然通过将每个样本的上述结果相乘而得到。这是一个我们可以处理为 GLM 问题的指数族分布形式吗?其 PMF 可以重新表达为:

$$f(\mathbf{y}_i \mid n_i, \boldsymbol{\pi}_i) = \exp\left[\underbrace{\left(\frac{\mathbf{y}_{i1}}{n_i}, \frac{\mathbf{y}_{i2}}{n_i}\right)}_{y'_1} \cdot \underbrace{\left(\log\left(\frac{\boldsymbol{\pi}_{i1}}{1-\boldsymbol{\pi}_{i1}-\boldsymbol{\pi}_{i2}}\right), \log\left(\frac{\boldsymbol{\pi}_{i2}}{1-\boldsymbol{\pi}_{i1}-\boldsymbol{\pi}_{i2}}\right)\right)'}_{\theta_1} - \underbrace{(-\log(1-\boldsymbol{\pi}_{i1}-\boldsymbol{\pi}_{i2}))n_i}_{b(\theta_2)} + \underbrace{\log\left(\frac{n_i}{y_{i1}!\ y_{i2}!\ (n-y_{i1}-y_{i2})!}\right)}_{c(y)} \right]$$

其中的 n_i 是第 i 个个案的权数。这明显是一个指数族形式,虽然 $\mathbf{y}'_i$ 和 $\boldsymbol{\theta}_i$ 都有一个两维结构。从上式中得到的两维连接函数为:

$$\begin{aligned}\boldsymbol{\theta}_i &= g(\boldsymbol{\pi}_{i1}, \boldsymbol{\pi}_{i2}) \\ &= \left(\log\left(\frac{\boldsymbol{\pi}_{i1}}{1-\boldsymbol{\pi}_{i1}-\boldsymbol{\pi}_{i2}}\right), \log\left(\frac{\boldsymbol{\pi}_{i2}}{1-\boldsymbol{\pi}_{i1}-\boldsymbol{\pi}_{i2}}\right)\right)\end{aligned}$$

因此,我们可以用以下方式来解释某一个被访者的结果:

$$\begin{aligned}&\log\left[\frac{P(\text{事件 }1)}{P(\text{参照组})}\right] \\ &= \log\left[\frac{\boldsymbol{\pi}_{i1}}{1-\boldsymbol{\pi}_{i1}-\boldsymbol{\pi}_{i2}}\right] \\ &= \mathbf{X}_i\boldsymbol{\beta}_1 \\ &\log\left[\frac{P(\text{事件 }2)}{P(\text{参照组})}\right] \\ &= \log\left[\frac{\boldsymbol{\pi}_{i2}}{1-\boldsymbol{\pi}_{i1}-\boldsymbol{\pi}_{i2}}\right] \\ &= \mathbf{X}_i\boldsymbol{\beta}_2\end{aligned}$$

这个模型的估计结果显示在表 3 中。从结果中可以看出,没有证据表明存在性别效应(与其他研究结果不同),但是相对于上学年数少于 12 年的参照组来说,有较强证据表明其他三组都更为幸福。有趣的是,随着教育年数增加,非常幸福与不幸福的区分在增加,但是较幸福与不幸福的区分却没有增加。偏差残差为 8.68,在 6 个自由度下这是非常好的拟合了(不在卡方分布的尾部),因此它比饱和模型有改善。

表 3　三类别的多项模型结果

	截距	女性	上学年数 12	上学年数 13~16	上学年数 17 以上
较幸福	1.129	−0.181	0.736	1.036	0.882
	(0.148)	(0.168)	(0.196)	(0.238)	(0.453)
很幸福	0.474	0.055	0.878	1.114	1.451
	(0.161)	(0.177)	(0.206)	(0.249)	(0.455)

——Jeff Gill

(高勇译校)

参考文献

Fisher, R. A. (1925). Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society*, *22*, 700-725.

Fisher, R. A. (1934). Two new properties of mathematical likelihood. *Proceedings of the Royal Society of London*, *Series A*, *144*, 285-307.

Gill, J. (2000). *Generalized linear models*: *A unified approach*. Thousand Oaks, CA: Sage.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). New York: Chapman & Hall.

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, *Series A*, *135*, 370-385.

几何分布(Geometric Distribution)

几何分布用来为一个二项分布取得第一次成功(r)前的失败次数($n-1$)建模。它是负二项分布【Negative Binomial Distribution】或 Pascal 分布【Pascal Distribution】在 $r=1$ 时的特例。几何分布是一系列的二项试验。由于二项试验具有独立性,希望出现的试验结果可以相乘来决定第一次成功发生的总概率【Probability】。例如,几何分布可以为掷硬币得到第一次“反面”所需要的次数建模。

如果一个离散的随机变量【Random Variable】X 的概率质量函数形式如下,则称其服从几何分布:

$$P(X=x)=p(1-p)^{(x-1)} \qquad (1)$$

其中,

$$x=1,2,3,\cdots \qquad (2)$$

p=成功概率,

$$0<p<1 \qquad (3)$$

如果 X 服从几何分布,每次试验必须只有两种可能结果:成功与失败。每次试验的结果必须是统计独立的,所有试验必须在有限的试验数范围内有相同的成功概率。

如果 X 是服从几何分布的一个随机变量,那么其期望为:

$$E(x)=\frac{1}{p} \qquad (4)$$

方差为:

$$\mathrm{Var}(x)=\frac{1-p}{p^2} \qquad (5)$$

几何分布有一个特性,它是“没有记忆的(memoryless)”。这意味着,观察到连续几次失败的概率,并不取决于之前已经进行了多少次试验。所以,已经观察到 100 次失败之后接下来五次试验都是失败的可能性,与你只观察到 10 次失败之后的可能性是一样的。分布“忘记”了已经发生了多少次失败。

几何分布可以和指数分布相类比。几何分布是“离散”的,因为你只可能得到整数的掷硬币次数(比如说,你不可以掷 1.34 次硬币)。指数分布是连续的。它是得到第一次成功前所需要的等候时间,而不是成功前的失败次数。

——Ryan Bakker

(高勇译校)

*也可参见分布【Distribution】。

参考文献

Johnson, N. L., & Kotz, S. (1969). *Discrete distributions*. Boston: Houghton Mifflin.

King, G. (1998). *Unifying political methodology: The likelihood theory of statistical inference*. Ann Arbor: University of Michigan Press.

Gibbs 抽样(Gibbs Sampling)

Gibbs 抽样是一种迭代蒙特卡洛程序,它已经被认为是用于近似复杂概率分布的

非常有效的统计工具,不用 Gibbs 抽样就不能构建这些复杂概率分布。这种技术是被称为马尔科夫链蒙特卡洛法【Markov Chain Monte Carlo, MCMC】的一组普遍程序的特例。Gibbs 抽样最重要的应用之一就是由于存在缺失值而变得复杂化的贝叶斯推论问题。社会科学中的其他应用还有层次线性模型分析和项目反应模型分析。对于 MCMC 方法背后的理论感兴趣的读者,可以在格尔曼、卡林、施特恩和鲁宾(Gelman, Carlin, Stern, & Rubin, 1995)和谢弗(Schafer, 1997)的著作中找到非常好的评述。

有一个很简单的例子可以很好地解释 Gibbs 抽样。假定我们知道两个变量 x_1 和 x_2的联合概率分布的统计形式。再假定我们对于每个变量的边缘分布感兴趣,和/或对于两个变量的某个函数(如 $x_1 - x_2$)的概率分布感兴趣。如果联合分布的数学形式并不复杂(如两个变量服从二元正态分布),我们就可以用分析方法构建出希望知晓的概率分布。但是,假定联合分布并不标准,因此不可能应用分析程序。在这种情况下我们应该怎么处理?如果我们可以从联合分布中抽取样本量为 N 的蒙特卡洛模拟样本,就能够构建经验相对频率分布(empirical relative frequency distributions),来估计我们感兴趣的分布。但是,假定由于分布的复杂性,从联合分布中抽样也非常困难,这时,就要用到 Gibbs 抽样了。在许多问题中,即使我们不能从联合分布中抽样,我们却能够很容易地从满条件分布(full conditional distributions)中,如 $P(x_1 \mid x_2)$ 和 $P(x_2 \mid x_1)$ 构建蒙特卡洛抽样方法。Gibbs 抽样采用如下迭代方案:给定初始值 x_1,从 $P(x_2 \mid x_1)$ 中抽取一个 x_2的值。给定抽取出的 x_2值,从 $P(x_1 \mid x_2)$ 中抽取一个新的 x_1的值。如果我们继续这一迭代方案通过了“充分”长的考验期(burn-in period),那么在这之后得到的所有(x_1, x_2)都可以视为从我们希望知晓的二元分布中得到的样本。通过分别对 x_1, x_2和$x_1 - x_2$构建相对频率分布,我们可以近似出希望知晓的分布。

在对不完整数据的贝叶斯分析中,面临的问题是只能用观察数据(observed data, OD)去构建未知参数(unknown parameter, UP)的后验分布。如果存在缺失数据(missing data, MD),后验分布的形式会极其复杂。但是,如果我们令 $x_1 = \mathrm{UP}$, $x_2 = \mathrm{M}$,条件为 O,我们就可以很容易地使用 Gibbs 抽样方案近似出希望知晓的后验分布和各个边缘分布。

Gibbs 抽样使用中尚未解决的问题是,如何确定考验期的长度以确保收敛,以及如何在适宜的考验期给定时构建样本。

BUGS 软件包是一种功能强大而且相对操作简便的程序(及参考资料),可以使用 Gibbs 抽样来运行许多种类的统计分析,软件可以从 www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml 免费下载。

——Alan L. Gross

(高勇译校)

参考文献

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. London: Chapman & Hall.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.

基尼系数(Gini Coefficient)

基尼系数是对某一总体中财富不平等或收入不平等的测量,取值为0~1。这个系数是对从整个总体中收入或财富的洛伦兹分布得到的图的测量(参见不平等测量【Inequality Measurement】)。这个图画出了"拥有财富或收入的百分比"和"总人口中的百分位数"的关系。分布完全平等将会产生一条45°的直线。普通的分布会有图1所示的形状。基层系数就是这个曲线与45°线之间的面积和45°线与x轴之间的面积之比。

——Daniel Little

(高勇译校)

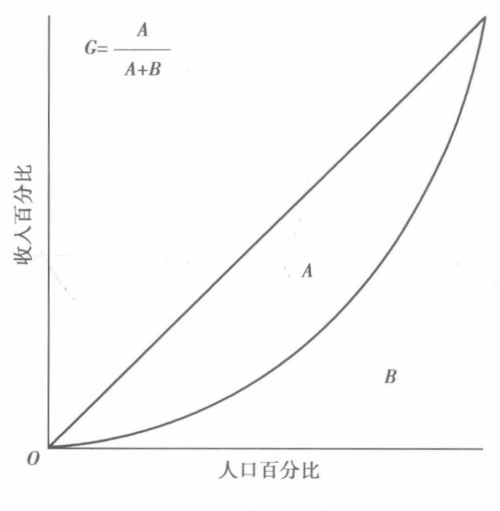


图1 收入的洛伦兹分布

GLIM

GLIM是一种用于估计广义线性模型【Generalized Linear Models】的统计软件包。它在处理用户自定义的宏时非常灵活。更多的信息,请参见软件网址 www.nag.co.uk/stats/GDGE_soft.asp①。

——Tim Futing Liao

(高勇译校)

土著化(Going Native)

在研究方法光谱的更为定性的一端,"土著化"一词概括的是研究者与"被研究者"之间逐渐发展出了一种"极其融洽(overrapport)"的感觉,到了研究者本质上已经"成为了"被研究者中一员的程度。帝国和殖民地统治精英给这个词强加了过分接纳"土著"文化和生活方式的负面含义,但其起源常常被归功于马林诺夫斯基,他倡议人类学家参与到所观察的文化中,以便增加对这些文化与人民的了解(Kanuha,2000)。这种对参与的倡议与更占主导地位的期望(社会)研究者形成客观性【Objectivity】的化身形成了截然相反的对比。"忠诚"于群体,进入群体内部(inside),与群体一起共事,可能会导致对这个群体及其行动完全采取一种没有批判、没

① GLIM1993年发布最后版本GLIM4后,已经不再更新版本。此网址目前也已失效。——译者注

有质疑的立场；研究的“标准”可能会被质疑，其效度【Validity】也受到威胁（Stanley & Wise, 1993）。研究中“作为个人的研究者”的出现，会被人认为研究者明显没有能力与所参加的事情保持“合适”的距离，最终损害“作为学术的研究者”的声音的权威性。但是，如卡努哈（Kanuha, 2000）所注意到的，这种看法已经受到越来越多的挑战，因为“曾经局限于白种男性异性恋研究者的排外的学术领域，已经融入了有色人种、来自西班牙语区的人，他们的祖先也许就是他们的导师和现在同事的土著研究对象”。（p.440）

从“边缘”视角来写作，研究者的角色——“我们”与“他们”的关系以及研究者和“被研究社区”的关系——在近年来已经越来越被解构和问题化，最终导致人们越来越意识到对于任何研究者的立场性（positionality）和情境性（situatedness）的意义有必要进行批判性和反思性的思考。在此关键性的概念是“私人的即政治的”（Stanley & Wise, 1993）及其反命题（Fuller, 1999），这说明了对于许多研究者而言，其私人的、政治的、学术的以及很多时候社会活动的身份都是同一的——许多研究者在他们进行的研究中一直是或已经是其中的一分子了。因此，在他们占据的空间中，其知识的情境性及其立场性不断地被重新界定，不断进行批判性处理，在此空间中必然要求去除研究者、社会活动者、老师、私人这些范畴之间的人为边界，要求在这些不同身份间不断移动以便投入这些身份之间和之中（Routledge, 1996）。在此，除了研究的地理定位、学科定位、政治立场和人格特点，还会对研究者的社会定位（从阶层、性别、种族等角度出发）进行不断提问（Routledge, 1994）。除此之外，对所作研究的任何评论都应包括此类提问，使之透明并作为研究过程的有机而不可或缺的部分得到彰显，借此避免将研究者（及研究本身）错误而具误导性地展现为被动的、超然的、中立的。

——Duncan Fuller

（高勇译校）

参考文献

Fuller, D. (1999). Part of the action, or "going native"? Learning to cope with the politics of integration. *Area*, *31*(3), 221-227.

Kanuha, V. K. (2000). "Being" native versus "going native": Conducting social work research as an insider. *Social Work*, *45*(5), 439-447.

Robinson, J. (1994). White women researching/representing "others": From anti-apartheid to postcolonialism. In A. Blunt & G. Rose (Eds.), *Writing, women and space* (pp. 97-226). New York: Guilford.

Routledge, P. (1996). The third space as critical engagement. *Antipode*, *28*(4), 399-419.

Stanley, L., & Wise, S. (1993). *Breaking out again: Feminist ontology and epistemology*. London: Routledge.

拟合优度量度（Goodness-of-Fit Measures）

拟合优度量度是从数据样本【Sample】中计算出的一个统计量【Statistics】，它测量的是样本数据与所考虑的模型【Model】的一致程度。在拟合优度量度中，R 方【R-Square】（表示为 R^2）的知晓度可能最高，在此我们就把它作为入手的示例。在线性回归【Linear Regression】分析中，变量【Variable】Y 用一个模型来近似，它被设定

等于 k 个解释变量【Explanatory Variable】的线性组合值 $\hat{Y}$，$\hat{Y}=\hat{\beta}_0+\hat{\beta}_1X_1+\hat{\beta}_2X_2+\cdots+\hat{\beta}_kX_k$。模型中的系数【Coefficient】，$\hat{\beta}_0$，$\hat{\beta}_1$，$\hat{\beta}_2$，$\cdots$，$\hat{\beta}_k$，需要从数据中进行估计，以使得 Y 与 $\hat{Y}$ 差值的平方和最小，Y 与 $\hat{Y}$ 差值的平方和也被称为误差平方和【Sum of Squared Errors，SSE】，$\mathrm{SSE}=\sum(Y-\hat{Y})^2$。$R^2$ 测量的是可观察变量 Y 与回归模型 $\hat{Y}$ 之间的拟合优度，定义为 $R^2=1-\frac{\mathrm{SSE}}{\mathrm{SST}}$，其中的 $\mathrm{SST}=\sum(Y-\bar{Y})^2$，表示围绕样本均值 $\bar{Y}$ 的偏差平方和。它的取值范围为 0~1，0 表示完全不能拟合，1 表示完美拟合。它具有普遍吸引力的原因在于，它能够解释为 Y 与 $\hat{Y}$ 的相关【Correlation】的平方，或者模型能够解释的 Y 的总方差比例。

其他示例

拟合优度量度并不限于回归模型，而是可以适用于对研究总体【Population】的任何假设性陈述。以下三个示例显示了不同的模型和相应的拟合优度量度。

变量 X 在总体中服从正态分布，这一假设就是一个模型的例子。对 X 进行观察得到一个样本，拟合优度统计量 D 测量的就是样本数据在多大程度上遵循正态分布。此外，D 统计量可以用于 Kolmogorov-Smirnov 检验【Kolmogorov-Smirnov Test】来对 X 在总体中服从正态分布进行正规检验。

另一个模型示例是假设 X 分布的均值（表示为 μ）等于 100，即 $\mu_X=100$。从数据样本中可以计算出 t 统计量【t-Statistic】，测量的就是样本均值 $\bar{X}$ 与假设的总体均值 μ_X 之间的拟合优度。然后可以用 t 统计量来进行 t 检验，以决定基于样本数据是否可以拒斥假设 $\mu_X=100$。

又一个示例是独立【Independence】模型，即假设两个类别变量在总体中无关。如果独立模型成立，样本数据中得到的列联表【Contingency Table】的每个类别的期望数值被称为期望频数【Expected Frequency】，样本中实际观察到的数字被称为观察频次【Observed Frequency】。假设的独立模型与样本数据之间的拟合优度用卡方统计量【Chi-Square Test】χ^2 来测量，然后就可以用于 χ^2 检验来决定是否可以拒斥独立模型。

假设检验

上面的示例表明存在两种不同的拟合优度测量，分别用于不同的目标。每一种都概括了样本数据与模型间的一致程度，但是 R^2 是描述性统计量，可以对其进行清晰的直观解释，而 D，t，χ^2 是推论统计学【Inferential Statistics】，其内在意义更不明确。推论统计学的价值在于能够对拟合优度进行假设检验【Hypothesis Testing】，因为每个统计量的抽样分布【Sampling Distribution】是已知的，通常分布的名称与统计量的名称一致（如 t 分布、χ^2 分布）。即使模型准确地再现了所研究的总体，从样本数据中得到的拟合优度统计量也很少会表明数据与模型是完美拟合的，与完美拟合的背离程度因样本而异。这种变异体现于拟合优度测量的抽样分布中，如 D 分布、t 分布、χ^2 分布。这些拟合优度统计量只有在与其抽样分布相比较时才可以理解，因为如果样本统计量位于其抽样分布的尾部足够远处，就表明如果模型为真的于是现有的抽样数据的可能性很小（也就是说数据与模型间的拟合很差）。根据检验的拒斥标准和所选择的显著性水平【Significance Level】，样本统计量可以拒斥模型。

回归模型

一种拟合优度统计量通常适用于某一种具体模型。但是,在诸如线性回归模型的一组模型中,研究者通常用多种测量来评估拟合优度。下面的例子就是在回归分析中使用多种拟合优度量度。这些测量可以分为两大类:(a)概括模型的整体拟合度的标量测量;(b)比较两个备择模型的拟合优度的模型比较测量。有些测量只适用于线性回归模型,另外一些测量适用于用最大似然估计【Maximum Likelihood Estimation,MLE】进行估计的更广义的回归模型。后者包括线性回归模型、logistic回归模型,以及因变量为类别变量和有限变量的其他模型。

标量测量

标量测量用一个数字来概括模型的整体拟合度。R^2是此类测量的最佳示例。但是,R^2也有局限,有其他一些测量就是测量来克服这些局限之处的。R^2的一个缺点是它只适用于线性模型。在非线性模型或者类别结果的模型中,不存在与R^2完全对应的统计量,虽然在这些情况下存在几种伪R^2【Pseudo-R^2】测量来模拟R^2的性质。一种常见的伪R^2测量是最大似然R^2(用R^2_{ML}表示)。R^2_{ML}在用于线性回归模型时可以完全再现R^2值,但是它的用处更广泛,可以用于MLE的所有模型。要评估的模型M_k中有k个解释变量,此时存在两个统计量L_{M_k}和L_{M_0},L_{M_k}为模型M_k的估计参数下观察到样本数据的似然值;L_{M_0}为没有解释变量的零模型条件下观察到样本数据的似然值。R^2_{ML}是用两个似然值之比计算出来的,$R^2_{ML}=1-\left[\frac{L_{M_0}}{L_{M_k}}\right]^{\frac{2}{n}}$。对于类别结果来说,更直观的替代统计量为计数$R^2$(表示为$R^2_{count}$),即用模型能够正确分类的样本观察比例。对于替代统计量的更多讨论,可以参看Long(1997, pp. 102-109)和Menard(1995, pp. 22-37)。

R^2的第二个局限是它在模型中加入更多变量时不能变得更小。这就容易误导人认为变量越多的模型就是拟合更好的模型。调整的R^2【Adjusted R^2】(表示为R^2_{adj})是一种更好的替代统计量,它在模型中加入更多变量时并不一定会增大。它的计算方式是,如果与观察数相比包括了太多变量时就降低R^2的值,$R^2_{adj}=1-\frac{n-1}{n-p}(1-R^2)$,其中的$p$为参数数目,$n$为观察数。

赤池信息准则(Akaike's information criterion, AIC)是另一种常用的标量拟合测量。和R^2_{adj}类似,它也考虑到了样本中的观察数和模型中的变量数,更鼓励那些与观察数相比参数数目更少的模型。对于一个有着n个观察、p个参数的线性模型来说,AIC的定义为$\mathrm{AIC}=n\ln\left[\frac{\mathrm{SSE}}{n}\right]+2p$。一个更一般性的公式可以适用于用MLE估计出来的模型。如果L是模型成立时观察到样本数据的似然值,那么$\mathrm{AIC}=\frac{2\ln L+2p}{n}$。AIC可以解释为"每个观察"对模型的贡献,因此取值越小越好,它可以用于比较不同模型对同一样本数据的拟合,也可以用于比较同一模型对不同样本数据的拟合。

模型比较测量

标度测量除了可以提供拟合程度的概括测量外,还可以用于评估一个模型与另一个模型相比的整体拟合度。R^2、R^2_{adj}、某些伪R^2的增加值,都属于更一般的一种模型比较测量,它们都具有误差消减比例【Proportional Reduction of Error, PRE】的意义。当比较有着k个解释变量的多元线性模型M_k和没有解释变量的零模型M_0的拟

合程度时,可以用推断性的 F 统计量,也被称为F比率【F-Ratio】,这个统计量与 R^2 直接相关,在多数统计软件包中会自动报告。F 统计量比 R^2 更具优势,因为它是对模型拟合进行正规检验(即 F 检验)的基础,这种检验的虚无假设为多元模型 M_k 的拟合程度并不比零模型 M_0 有显著提升。这等效于检验模型中的所有斜率系数都等于 0 的假设,即 $\beta_1=\beta_2=\cdots=\beta_k=0$。

与线性模型中的 F 统计量类似,模型χ^2 统计量可适用于 MLE 的更广泛的模型。正如 F 统计量与 R^2有关一样,模型χ^2 统计量与 R^2_{ML}有关;正如 F 统计量是 F 检验的基础一样,模型χ^2 统计量是对虚无假设进行似然比【Likelihood Ratio】检验的基础,此时的虚无假设为多元 MLE 模型 M_k 的拟合程度并不比零模型 M_0 有显著提升,或者等效,$\beta_1=\beta_2=\cdots=\beta_k=0$。模型$\chi^2$ 统计量是如前所述的两个似然值(L_{M_k}和 L_{M_0})之比的函数,计算公式为$\chi^2=-2\ \ln\left[\frac{L_{M_0}}{L_{M_k}}\right]$。

F 检验和似然比检验的适用范围比上述示例中表明得更为广泛。它们可以用于检验任意两个模型间的拟合程度差异,只要其中一个模型中包括的变量是另一个模型中包括的变量的子集。它们可以检验一组变量或一个变量是否提升了拟合程度。对于 MLE 模型来说,除了似然比检验,还可以进行 Wald 检验(Wald test)或者拉格朗日乘数检验(Lagrange multiplier test)。在检验单个变量是否显著提升了模型拟合时,可以在线性模型和 MLE 模型中用 t 检验。

尽管这些检验可以用来对两个模型的拟合程度差异进行假设检验,但是要谨慎使用这些检验。因为拒斥虚无假设的概率会随着样本量增大而增加,在样本量较大时过分信赖这些检验会被误导,在变量效应实际上微不足道时却误以为变量显著地提升了拟合。贝叶斯信息量准则(Bayesian Information Criterion, BIC)是不受样本量影响的替代选择。BIC 可以用于许多测量模型提升的统计量,它是评估一个模型与另一个模型相比拟合提高程度的一种方法。例如,在应用于模型的χ^2 统计量来比较 M_k 和 M_0 时,$\text{BIC}_{M_k}-\text{BIC}_{M_0}=-\chi^2+k\ \ln\ n$。一般而言,在比较任何两个 MLE 模型 M_1 和 M_2 时,并不要求一个模型中包括的变量是另一个模型中包括的变量的子集,此时 $\text{BIC}_{M_1}-\text{BIC}_{M_2}\approx 2\ \ln\left[\frac{L_{M_2}}{L_{M_1}}\right]$。BIC 是更大的负值的模型更优,因此如果 $\text{BIC}_{M_1}-\text{BIC}_{M_2}<0$,则第一个模型更优。如果 $\text{BIC}_{M_1}-\text{BIC}_{M_2}>0$,则第二个模型更优。拉夫特里(Raftery, 1996)提出了基于两个 BIC 的绝对差值评估模型提升时的建议指导,详见表 1。有关 BIC 的更多讨论,请参看拉夫特里(Raftery, 1996, pp. 133-139)和朗(Long, 1997, pp. 110-112)。

表 1 拟合优度提升的证据强度

绝对差值,$\lvert \text{BIC}_{M_1}-\text{BIC}_{M_2} \rvert$	证据强度
0~2	弱
2~6	正面
6~10	强
>10	非常强

——Jani S. Little

(高勇译校)

参考文献

Greene, W. H. (2000). *Econometric analysis* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.

Menard, S. (1995). *Applied logistic regression analysis*. Thousand Oaks, CA: Sage.

Pampel, F. C. (2000). *Logistic regression: A primer*. Thousand Oaks, CA: Sage.

Raftery, A. E. (1996). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology* (Vol. 25, pp. 111-163). Oxford, UK: Basil Blackwell.

隶属度模型(Grade of Membership Models)

隶属度(GoM)模型将一组对象以模糊划分的形式进行分类,其中人可能部分地隶属于多个组。与离散划分的分类方法(如聚类分析【Cluster Analysis】、潜类分析【Latent Class Analysis, LCA】)相比,隶属度模型的优势在于它生成的类别更少,类别与测量变量的参数关系更易于解释。事实上,LCA 可以视为 GoM 模型在类别清晰、每个人只隶属于一个组时的特例(即不存在组内异质性)。在因子分析中,模型的构建只使用了二阶矩(即协方差矩阵或相关矩阵),所以它不能描述非正态多元分布,但是如果有多个不同组(小组)构成了分析的个案全集,就可能会呈现非正态多元分布。

描述

GoM 模型处理的是代表测量的一组随机变量$\{Y_{ij}\}_{ij}$。其中的 i 代表一组对象 $\mathcal{I}$(在人口学中,对象通常就是个人),j 代表一系列测量 $\mathcal{J}$,Y_{ij}的取值范围是一个有限集 $\mathcal{L}_j$,一般性地,这个集可以被视为一组自然数$\{1,\cdots,L_j\}$。每个个体 i 都与一个随机变量向量$(Y_{ij})_j$ 相关联,这个向量中包含了个体的所有可获得信息。这里的理念是,在随机向量$(Y_{ij})_j$ 的空间中构造一个凸空间或者"基准"$\{(Y^0_{kj})_j\}_k$,$k=1,\cdots,K$,对每个人产生其在这个"基准"中的位置系数$(g_{ik})_k$,以使方程 $\Pr(Y_{ij}=l)=\sum^K_{k=1}g_{ik}\Pr(Y^0_{kj}=l)$ 在约束条件 $g_{ik}\geqslant 0,\sum_k g_{ik}=1$ 下成立。可以将这些结果视为 K 个模糊集合,或者"完美面相(pure profiles)",$(Y^0_{kj})_j$ 描述了第 k 个"面相(profile)"中的一个代表性成员(对 k 而言 $g_{ik}=1$)。对"面相"的解释源于对隶属度模型产生的系数 $\lambda_{kjl}=\Pr(Y^0_{kj}=1)$ 的事后分析。

GoM 模型的算法【Algorithm】需要寻找使得似然值 $\prod_{i\in I}\prod_{j\in J}\prod_{l\in L_j}(\sum^K_{k=1}g_{ik}\lambda_{kjl})^{y_{ijl}}$ 最大的 g_{ik} 和 λ_{kjl},约束条件为 $g_{ik}\geqslant 0,\sum_k g_{ik}=1$,$\lambda_{kjl}\geqslant 0$,以及 $\sum_l\lambda_{kjl}=1$,其中如果对第 i 个个体的第 j 个测量得到结果 l,则 y_{ijl} 取值为 1,否则取值为 0。如果没有任何回答,那么所有的 $y_{ijl}=0$,变量缺失,因此从似然中删除此变量。λ_{kjl} 决定了 Y^0_{kj} 的分布——即 $\Pr(Y^0_{kj}=l)$——g_{ik}就是隶属度的估计。g_{ik}的似然估计是一致的(也就是说,当个体数目和测量数目趋向于无穷时,g_{ik}估计收敛于真值)。与在标准统计模型中一样,λ_{kjl} 在对 i 取平均值后也具有统计一致性。

示例

在对全国长期照料调查的分析中(Manton & XiLiang, 2003),研究者选择了 27 个对人的生活能力的不同测量,有些编码是二分的,有些编码是四分的。GoM 模型分析的结果辨析出了六个"面相"($K=6$),其位置是从 27 个可观察变量的角度用 6 个向量 λ_{kjl} 来定义的。在这个分析中,第一个面相是完全没有失能(no disability),第六个面相是很可能全面性地功能性失能,其他四种类型分成了两对向量,第一对描述的是日常生活中工具性活动的损伤程度,第二对描述的是日常生活中其他行为的损伤程度。由于 g_{ik} 中

有着显著变异,潜类分析中用6个组所能描述的变异没有隶属度模型中多。由于 g_{ik} 并不服从正态分布,因子分析也不能仅用6个维度来完全描述数据(包括高阶矩)。

历史与书目

隶属度模型是1974年首次提出的(Woodbury & Clive, 1974)。许多书籍(Manton & Stallard, 1988; Manton, Woodbury, & Tolley, 1994)都对这种方法作了详尽描述,并给出了应用示例。一些论文(Tolley & Manton, 1992)中也包含了对于 g_{ik} 和 λ_{kjl} 参数估计的重要统计性质的证明。

——Kenneth Manton
Mikhail Kovtun
(高勇译校)

参考文献

Manton, K. G., & Stallard, E. (1988). *Chronic disease modelling: Vol. 2. Mathematics in medicine*. New York: Oxford University Press.

Manton, K. G., & XiLiang, G. (2003). *Variation in disability decline and Medicare expenditures*. Working Monograph, Center for Demographic Studies.

Manton, K. G., Woodbury, M. A., & Tolley, H. D. (1994). *Statistical applications using fuzzy sets*. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley.

Tolley, H. D., Kovtun, M., Manton, K. G., & Akushevich, I. (2003). Statistical properties of grade of membership models for categorical data. Manuscript to be submitted to *Proceedings of National Academy of Sciences*, 2003.

Tolley, H. D., & Manton, K. G. (1992). Large sample properties of estimates of discrete grade of membership model. *Annals of the Institute of Statistical Mathematics*, *41*, 85-95.

Woodbury, M. A., & Clive, J. (1974). Clinical pure types as a fuzzy partition. *Journal of Cybernetics*, *4*(3), 111-121.

Granger 因果关系(Granger Causality)

计量经济学家克里弗·格兰杰(Granger,1969)对Granger因果关系【Granger Causality】给出了一个被广泛使用的定义。Granger因果关系这一概念的关键之处在于不存在即刻因果关系(也就是说,所有因果过程都有时间差)。如果加入 X_t 的历史信息,能够比只使用 Y_t 的历史信息更准确地预测 Y_t,那么我们称 X_t 是 Y_t 的"Granger原因"。Granger因果关系不是单维度的;也就是说,如果加入 Y_t 的历史信息,能够比只使用 X_t 的历史信息更准确地预测 X_t,那么我们同样可以称 Y_t 是 X_t 的"Granger原因"。

可以用双变量向量自回归模型(bivariate vector autoregression, VAR)来进行Granger因果检验。X 和 Y 的VAR模型设定中,每个变量都是自身的 k 个滞后与其他变量的 k 个滞后的函数,即

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_k Y_{t-k} + \lambda_1 X_{t-1} + \cdots + \lambda_k X_{t-k} + \varepsilon_t \quad (1)$$

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + \cdots + \alpha_k K_{t-k} + \gamma_1 Y_{t-1} + \cdots + \gamma_k Y_{t-k} + \zeta_t \quad (2)$$

可以用 F 检验或拉格朗日乘数(LM)检验来决定 X 是否是 Y 的"Granger原因",或者 Y 是否是 X 的"Granger原因"。例如,要用LM检验来看 X 是否是 Y 的"Granger原因",就需要在所有的 λ 都是零的假定下去估计式(1)。从这个回归中得到的残差对式(1)中的所有变量进行回归。从第二个回归中得到的 $R^2 \times T$(观察数)就是检验统计

量(服从卡方分布X^2,df=2k+1)。如果原假设被拒斥,则说明 X 是 Y 的"Granger 原因"。西姆斯(Sims, 1972)提出了上述检验的另一种 VAR 表达式。非平稳性对于上述检验的有效性是否有影响,对此是有争议的。Granger 因果关系检验也可以在 ARIMA 模型框架下进行(参看 Freeman, 1983)。

Granger 因果关系的概念已经被吸纳到亨德里等人提出的外生性定义中(参看 Charemza & Deadman, 1997, Chap.7)。注意,要确保 Y 的单方程模型中 X 的效应推论成立,并不需要确定 X 是 Y 的 Granger 原因。确定 X 弱外生(weakly exogenous)于 Y 就足够了(即在没有 X 的模型的信息时,可以有效地估计出 Y 的模型参数)。但是,如果希望对 Y 进行预测,就必须考察 X 是否强外生(strongly exogenous)于 Y。强外生性相当于弱外生性加上不存在 Granger 因果关系(即 Y_t 不影响时间点 t 或之后任何时点 $t+k$ 上的 X)。如果 Y 并不是强外生于 X,只要反馈作用发生在预测区间中,就必须考虑 Y 对 X 的反馈作用。

——Harold D. Clarke

(高勇译校)

参考文献

Charemza, W. W., & Deadman, D. F. (1997). *New directions in econometric practice* (2nd ed.). Cheltenham, UK: Edward Elgar.

Freeman, J. F. (1983). Granger causality and the time series analysis of political relationships. *American Journal of Political Science*, *27*, 325-355.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*, 24-36.

Sims, C. A. (1972). Money income and causality. *American Economic Review*, *62*, 540-552.

图形建模(Graphical Modeling)

图形模型是一个对一系列随机变量建立的参数统计模型。在这种模型下,变量间的相关可以用一个数学图形来表示。图形建模就是选择何种图形可以最佳地表示数据的过程。

统计学中利用数学图形可以上溯至1920年代和1930年代中休厄尔·赖特的路径图【Path Diagram】。但是,直到1980年约翰·达罗克、斯特芬·劳里岑和特里·斯皮德发表其开创性论文以后,一种具有明晰的概率解释的图形构建方法才被提了出来。这种图形自此之后被称为条件独立图(conditional independence graph),或者简称为独立图(independence graph)。

变量X_1和X_p之间的(条件)独立结构可以用一个有 p 个端点(或称为点)的数学图形来表示,这 p 个端点对应于 p 个变量。当且仅当X_i和X_j在其余 $p-2$ 个变量给定时条件独立,端点 i 和 j 之间的边(或称为线)可以缺失。例如,如果X_1到X_4满足如下条件独立:X_1在X_3和X_4给定时独立于X_2, X_1在X_2和X_4给定时独立于X_3,X_2在X_1和X_4给定时独立于X_3;那么如图 1 所示,它们的独立图有四个端点,端点对(1, 2),(1, 3)和(2, 3)之间的边是缺失的。

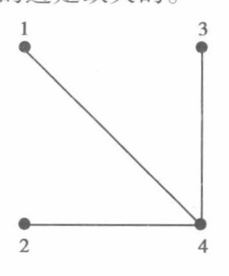


图 1 一个独立图

任何一组联合分布的随机变量都可以

有一个独立图,但是为了对给定数据构建图形,分析者需要一些模型来方便地将条件独立参数化。对于连续变量【Continuous Variable】而言,可以使用多元正态分布;对于离散变量【Discrete】,可以使用对数线性模型【Log-Linear Model】;对于离散和连续变量兼有的情况,可以使用条件高斯分布。离散和连续变量兼有时,在图中通常用实心点来表示离散变量,用空心圆来表示连续变量。

在搜寻拟合良好的图形模型时,通常首先计算去除一边后的偏差(deviances):从图中依次去掉每边,然后进行似然比统计量【Likelihood Ratio】检验。偏差被用来与事先选定的临界水平相比,以获得一组显著的边,这些边必须在图中保留。然后再考虑那些可能从图中去掉的不显著的边,继续使用类似于同归中选择最佳回归自变量子集的程序进行模型筛选。选择了一个拟合良好的图形后,可以从图中利用被称为马尔科夫特性的理论结果获取变量间关联的信息。

潜变量模型可以用包括代表不可观察变量的端点的图来表示。贝叶斯模型可以用包括代表参数的点来表示。端点之间是箭头而非直线的有向图,或者直线和箭头兼有的链图,可以用来表示变量间存在的(潜在)因果关系或者用来区分响应变量和解释变量。

——Peter W. F. Smith

(高勇译校)

参考文献

Darroch, J. N., Lauritzen, S. L., & Speed, T. P. (1980). Markov fields and log-linear models. *Annals of Mathematical Statistics*, *43*, 1470-1480.

Edwards, D. (2000). *Introduction to graphical modelling*(2nd ed.). New York: Springer-Verlag.

Whittaker, J. (1990). *Graphical models in applied multivariate statistics*. Chichester, UK: Wiley.

扎根理论(Grounded Theory)

扎根理论是指一种系统的归纳方法,用于进行以理论建构为目标的定性研究【Qualitative Research】。"扎根理论"指涉的内容包括两种:(a)一种包括多种灵活的方法策略的方法;(b)这种研究的成果。研究者越来越多地用这个术语来表示收集资料特别是分析资料的研究方法。扎根理论方法策略的目标是从资料分析中直接构建中层理论。其逻辑核心是这些方法具有的归纳性理论主旨。由此得到的分析结果的效力基于坚实的经验基础之上。这些分析可以产生能够解释所研究经验现象的集中的、抽象的、概念性的理论。

扎根理论相当重要,因为它(a)为如何进行定性研究提供了明确清晰、井然有序的指导原则;(b)为如何处理研究过程中的分析阶段提供了具体策略;(c)对于资料收集和资料分析工作进行了精简和整合;(d)推进了对定性资料进行概念分析;(e)使得定性研究成为一种正当合理的科学研究。扎根理论已经成为一种标准的社会研究方法,影响了诸多领域和专业的研究者。

扎根理论一直是一种被误解的方法,虽然许多研究者声称在使用它。许多定性研究者常常会宣称自己在进行扎根理论研究,却并没有真正理解和采纳其独特的指导原则。他们可能采用的只是一两种分析策略,或者将定性分析和扎根理论混淆了。相反地,其他一些研究者则是简单化地、机械地采用扎根理论方法。上述两种路数都不能够体现出扎根理论的特点:灵活但不零散的

研究方式，开放但坚持指导原则的分析，基于经验资料但要进行富于想象力的理论概括。结果，通过扎根理论方法来发展中程理论的潜力并没有充分实现。

扎根理论方法的发展

扎根理论的创始人是格拉泽和斯特劳斯（Glaser & Strauss, 1967），他们试图发展出一套系统的程序来分析定性资料。他们想对社会过程进行理论分析的构建，从而对社会过程进行抽象理解。他们对这种方法的第一次陈述是《扎根理论的发现》（*The Discovery of Grounded Theory*），雄辩地说明了定性研究是一种可信的方法，而不仅仅是发展定量测量前的初步了解，因此定性研究是正当合理的。当时，基于狭隘的实证主义【Positivism】，社会研究的主流是依赖于定量化的，定性研究的重要性因此被忽视。格拉泽和斯特劳斯提出系统的定性研究有其自身逻辑，可以产生理论。他们的著作对于复兴定性研究、继承芝加哥学派的民族志传统作出了贡献。他们试图使定性研究不仅是描述研究，而且要进入解释性理论框架的领域中，提供对研究现象抽象的、概念化的理解。对于他们来说，这种抽象理解不同于扶手椅上的逻辑演绎理论，因为扎根理论包含如下特征：严格基于资料、有用、言之有物、持久、可修正、具有解释力（Glaser, 1978, 1992; Glaser & Strauss, 1967）。

斯特劳斯继承了芝加哥学派的实用主义和符号互动论的学术传统，格拉泽在哥伦比亚大学保罗·拉扎斯菲尔德那里接受了严格的定量方法训练，扎根理论将这两者结合了起来。斯特劳斯为扎根理论引入了主体（agency）、凸生（emergence）、意义等概念，引入了对行动的实用主义研究。格拉泽运用其分析技能对定量分析进行编码。他们对研究社会过程（social processes）都有着浓厚兴趣。通过使用诸多方法策略来解释特定社会场景或特定经历（如患有慢性病）中的社会或社会心理过程，他们发展出了一种得到实质理论的途径。随后的扎根理论致力于理解所研究的过程，显示它产生和变化的原因及条件，解释其后果。格拉泽和斯特劳斯研究了其他实质领域中的许多类似过程，找寻在多种情境中的相关资料以改进他们正在形成的理论，在这一过程中他们将其方法进行了正规理论化。

虽然《扎根理论的发现》使方法争论为之改观，启发了几代定性研究者，但是格拉泽（Glaser, 1978）的著作《理论敏感性》（*Theoretical Sensitivity*）对于这一方法进行了最明确的早期阐述。自此之后，格拉泽和斯特劳斯在扎根理论上已经分道扬镳（Charmaz, 2000）。格拉泽延续了他早期对这一方法的阐释，依赖于直接而且常常是狭隘的经验主义【Empiricism】。斯特劳斯将方法向验证的方向推进，他和朱丽叶·科尔宾合作的著作（Strauss & Corbin, 1990, 1998）是这一方向的进一步发展。

斯特劳斯和科尔宾（Strauss & Corbin, 1990, 1998）的扎根理论版本支持的是他们的新方法，而不是早期扎根理论策略中的比较方法。格拉泽一如既往地研究基本社会过程，使用比较方法，构建理论范畴间的抽象关系。格拉泽的立场背后的假定是，独立于观察者的外在现实，中立的观察者，资料是被发现的。这两种路数都有着基于实证主义的强烈客观主义研究要素。但是，格拉泽（Glaser, 1992）认定斯特劳斯和科尔宾的程序把资料和分析塞到先入为主的范畴当中，因此违背了扎根理论的基本信条。格拉泽也认为研究参与者固然会告诉研究者在他们的情景中什么是最重要的，但是参与者常常会视基本的过程为理所当然。诠释性研究展示了扎根理论可以如何促进建构主义，进而淡化其实证主义色彩。

扎根理论已经获得了广泛赞誉,因为它的系统策略有助于整理和分析定类资料。它的唯实论【Realism】特点、严格性、实证主义假定使得定量研究者也能够接受它。它的灵活性和正当性吸引了持有不同理论兴趣和实质兴趣的定性研究者。

使用扎根理论方法

扎根理论方法提供了处理资料的指导原则,能够促进对资料的思考方式。这些方法本质上是比较。通过资料与资料的比较、资料与理论范畴的比较、范畴与范畴的比较,研究者获得了新的想法(Glaser, 1992)。扎根理论中的发现是在研究者深刻的经验知识中、在研究者与资料的分析性互动中涌现出来的;发现并不是资料所固有的。虽然研究者的学科视角使得他们对于概念议题很敏感,但是扎根理论促使研究者以这些视角为起点而非终点。

这些方法需要研究者从资料收集开始就进行前期的资料分析工作。前期的分析工作又会指导扎根理论研究者的下一步资料收集。扎根理论策略要求研究者一直积极地投身于分析步骤当中。这些策略也要求进行检查以确保研究者正在形成的理论扎根于其力图去解释的资料。研究者收集哪些资料,如何处理资料,都会形塑他们对资料的看法。反过来,他们对资料的看法,包括他们对资料提出的问题,又会指引他们以后的方法步骤。此外,扎根理论方法不仅形塑了资料的内容,也决定了分析的形式。

扎根理论使得研究者对其工作采取一种渐进的分析立场和方法立场。他们在资料收集过程中逐步聚焦,最终聚焦到回答特定的分析问题上,聚焦到填补正在形成的分析中的空缺。因此,他们找寻的资料能够使他们对研究现象有全面的看法。扎根理论策略有助于对具体过程收集丰富的资料,而不是社会情境的普遍结构。这些策略也有助于逐步将分析更为聚焦、更为抽象。然后,扎根理论分析可以得到对所研究现象的有力解释,其他研究者从中可以演绎各种假设。由于扎根理论的目标是产生理论,正在形成的分析会包括抽象的概念及其相互关系。在这一方法的所有支派中,下述扎根理论策略都是其逻辑核心。

资料编码

分析过程始于编码的两个阶段:开放编码或初始编码(open or initial coding),选择性编码或聚焦性编码(selective or focused coding)。在开放编码中,研究者分析性地处理资料,在其中发现他们所见到的。他们仔细查看资料,对每一行文本进行分析和编码(示例可参看表 1)。

这种逐行编码让研究者能够一直对新想法抱有开放态度,避免了将资料塞到先入为主的范畴中的倾向。这样积极地投入到分析中去,有助于发展出理论敏感性,因为研究者开始对资料提出分析性问题。扎根理论家的基本问题是:这里发生了什么?但是他们不仅是去描述行动,而且要确定行动的阶段、前提条件、性质和目的。格拉泽和斯特劳斯的早期陈述明确说明分析的是在情境中的基本社会过程或社会心理过程。对基本过程的研究使得分析是有说服力的。但是,多个重要过程可能是同时进行的:研究者可以同时研究几个过程,展示它们的相互关联。在更大的意义上来说,扎根理论有助于将任何题目都视为一种过程来研究。

初始编码强迫研究者去思考资料的细节,对它们进行阐释——当下阐释。如表 1 中所显示的,扎根理论编码应当是积极的、具体的、简短的。

表 1 初始编码示例

逐行编码	一位患有慢性病的 68 岁老年女性的访谈记录
	我们（一位患有多发性硬化症的朋友）
互助	对各自来说，有点像对方的支持者。
倒霉	在她倒霉不顺利的时候，或者
不允许自怜	我们都觉得“我好可怜”的时候，你知道，
彼此发令	“滚开点”！你知道，我们之间可以不讲道理，
	也都能理解这一点。但是对于弗雷德
	（患有糖尿病和心脏病的丈夫）来说，
与其他支持者相比较	无论我想干什么他总会支持我。
定义支持角色	……因为有他，我努力不老是想自己的
考虑他人	事情，……因为他会担心。如果他看
解释未言之意	我的眼神不太正常，我能理解那些眼
涉及对健康/患病的测量	神，你知道，就是说“她还好吗？”这样
避免绝望	他让我振作起来，我们对彼此都很好，
改变伤痛	因为我们彼此支持，我们需要对方。
承认他人非常重要	没有他，我不能想象怎么活。我知道
	我能活下去，但你知道，我宁愿死。

通过编码，研究者对资料有了新的视角，从多个角度来审视它。逐行编码弱化了那些研究者和研究参与者都会视为理所当然的假定，促使研究者将其中的意义、行动、言辞视为可以引发问题的研究对象。对资料的仔细查看为比较资料中的各个部分做好了准备。这样，研究者去比较同一个人的陈述或行动、比较不同的事件、比较不同参与者的经验、比较在不同时点的相似事件。例如，在表 1 中的被访者提到她的朋友和丈夫提供了不同形式的支持。研究者可以沿着其陈述中的这一线索，对支持及其产生效果的条件进行比较性分析。

在进行逐行编码时，研究者研究其资料，在看似不相关的资料间建立起联系，了解基本过程。然后他们会在接下来的资料收集中探索最为突出的过程。逐行编码在访谈和会议的转录中作用尤其显著。对民族志材料，它可能作用不大，这取决于实地笔记描述的层次和质量。因此，研究者可以对更大的单元进行编码，如段落或者对传闻或事件的描述。

初始编码引导研究者进入选择性编码或聚焦性编码。在对初始编码考察完之后，研究者要找到其中最频繁出现的编码，以及能够提供最为关键的概念切入点的编码。研究者用这些编码进行聚焦性编码。聚焦性编码是进行如下工作的分析工具：处理大量资料；重新评估前期资料中的内含意义、陈述和行动；随后产生正在生成的理论中的范畴。研究者越来越多地采用计算机辅助程序来更有效地进行资料编码和资料整理。这类程序是否会导致简化而机械的分析，这仍然是一个有争议的问题。

写备忘录（memo）

写备忘录是编码和报告初稿写作之间的关键性中间阶段。通过写备忘录，研究者以叙事形式发展其分析，从描述性材料推进到分析性陈述。在研究的前期阶段，备忘录写作包括考察初始编码，定义编码，

分解编码，找寻编码中内含的假定、意义和行动，找寻编码之间的关系。前期的备忘录有助于对直觉进行讨论，注意到含糊不清之处，提出问题，澄清思想，进行资料间的比较。

这些备忘录也会推动分析，因为它们促使研究者解释资料间的哪种比较是最具说服力的，何以如此，为何如此。这些比较有助于研究者定义参与者的世界和行动中最为重要的方面，并将其视为一种过程。随后，他们可以决定哪些过程是最重要的，然后在备忘录中探索和解释这些过程。写备忘录成了一种研究资料的方式。然后，研究者可以通过进一步的数据收集和分析，去追查资料中最有趣的线索。

写备忘录使得研究者可以定义和界定他们正在生成的理论范畴。然后他们继续写备忘录来发展这些范畴，确定它们如何组合在一起。他们比较资料和相关范畴，比较范畴和范畴。在研究中逐步地写作更具理论性的备忘录，可以增加研究者解释资料的分析能力和进行诠释时的信心。

通过使备忘录逐渐更具分析性和涵盖每一个范畴，研究者明白了如何把这些范畴纳入所研究的过程中。他们也因此尽可能地克服了一个主要困难：如何将分析整合成为一个连续一致的框架。

理论性抽样

研究者在写备忘录解释其编码和范畴时，他们也会觉察到存在一些缺口，需要进行进一步的资料收集。随后，他们会致力于理论性抽样，这意味着收集更多的具体资料来说明、扩展、改进理论范畴，使这些范畴更为精确。因此，指导这种抽样的是研究者正在生成的理论，而不是对总体特征成比例的代表性。巧妙地使用这种抽样，研究者能够填充其正在生成的理论范畴中的缺口，回答尚未解决的问题，澄清何种条件下范畴能够成立，解释范畴的后果。

理论性抽样既可以在同一实地环境中进行，也可以在不同实地环境中进行。例如，研究者可以回到实地，提出更深刻的研究问题或者不同的研究问题来寻找新的资料；他也可以寻找新的参与者或具有类似特征的新环境，来获取新的看法。但是，如果研究者正在生成的理论要解释的是一个普遍性过程（generic process），他们就可以在不同的实质领域中研究他们的概念，发展出对这一普遍性过程的形式理论。通过理论性抽样，研究者使得其范畴更为深刻，其备忘录更为有用，其研究更为深入地扎根于资料当中。

构建和联系理论范畴

编码、写备忘录以及理论性抽样逐步引导研究者构建出能够解释所研究过程或现象的理论范畴。有些范畴是从开放编码中直接生成的，如研究参与者的讲述。在研究者的分析变得更为抽象，纳入更多的编码时，他也会不断想出其他范畴来。

扎根理论将核心范畴的饱和作为结束资料收集的标准。但是核心范畴的饱和是什么意思？应当采用谁对饱和的定义？如果范畴是具体而有限的，那么饱和很快就会达到。一些自我标榜为扎根理论的研究既不是扎根性的，也不是理论性的；他们的范畴缺乏解释效力，范畴间的关系要么根本不重要，要么早已为人所知。

对于有解释效力的扎根理论来说，其理论范畴应当是抽象的、清晰的、与其他范畴相整合的。因此，扎根理论家会完成如下任务：确定范畴起重要作用的背景（context）；定义每个范畴；界定每个范畴的基本特性；澄清范畴存在或变化的条件；指明何时何地某范畴如何与其他范畴相关联；确定这些关

系的后果。这些分析努力的成果是对于所研究过程有崭新的理论理解。

——Kathy Charmaz

（高勇译校）

参考文献

Charmaz, K. (2000). Grounded theory: Constructivist and objectivist methods. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 509-535). Thousand Oaks, CA: Sage.

Glaser, B. (1978). *Theoretical sensitivity*. Mill Valley, CA: Sociology Press.

Glaser, B. (1992). *Emergence vs. forcing: Basics of grounded theory analysis*. Mill Valley, CA: Sociology Press.

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory*. Chicago: Aldine.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Grounded theory procedures and techniques* (2nd ed.). Thousand Oaks, CA: Sage.

小组访谈(Group Interview)

小组访谈是指在正式或非正式场景中同时对多个人进行系统的提问(Fontana & Frey, 1994)。这种技术在社会科学中使用有限,在社会科学中主要强调的是对个人进行访谈。小组访谈最流行的用法是焦点小组法【Focus Group】,但是这种焦点小组法技术也有其他的形式,特别是在定性田野工作中。取决于研究目标,访谈者/研究者可以在正式或自然发生的场景中以结构方式或非结构方式来引导小组互动和进行回应。小组可以用于在探索性研究中,检验一个研究想法、一个调查问题或者一种测量技术。小组头脑风暴法还可以用来获得对一种特定行为的最佳解释。小组讨论也可以使研究者确定关键被访者(Frey & Fontana, 1991)。这种技术也可以与其他资料收集策略结合起来,来获得对同一研究问题的多种看法。小组访谈的另一个维度是,对于超越个体诠释的凸生意义进行现象学评估。

小组访谈可以像在头脑风暴或探索性研究中那样只有很有限的结构(即来自访谈者的指导),也可以像在团体列名法(nominal groups)或德尔菲法或多数市场焦点小组法中那些有相当严密的结构。被访者可以被带到一个正式场景中,如研究中心的实验室中,也可以在一个非正式的田野场景中模拟进行,如在街角、娱乐中心或邻里活动区。任何小组访谈,无论场景如何,都需要访谈者具有相当能力去捕捉到访谈机会,提出能够激发出所需回应的问题,得到有代表性的回应,而不是只得到那些只代表少数强势者的观点而忽略掉其他羞怯者的视角。访谈者必须在获取与研究目标有关的回应和获取凸生的群体回应之间进行平衡。因此,访谈者可以使用事先确定的问题清单来严格控制提问过程,也可以对回应的方向和内容不进行任何控制,只是让交谈和回应按其轨迹自由进行。

如果找不到合适的小组访谈场景,或者这样一个场景(如公共建筑、街角)并不适合进行实话实说的、常常有着唇枪舌剑的讨论,进行小组访谈就会存在问题。也有可能小组并不接纳研究者,无论是作为小组一员还是作为研究者。小组成员可能在观点上倾向于从众而不是多样性的,这样就捕捉不到回应的潜在差异性,因此很

难谈论敏感话题(Frey & Fontana, 1991)。最后,存在很高比例的无关资料;记录与编辑这些现场笔记可能困难重重,特别是在自然的非正式场景中;被访者在小组中更有可能是故作姿态。

优点是效率,在同一场景下可以同时接触到多个被访者。小组访谈提供了理解事件或行为的另外一种丰富维度;被访者可能被小组激发从而回忆起可能已经忘却的事情,研究者可能会洞察到与特定田野场景有关的社会关系。

——James H. Frey

(高勇译校)

参考文献

Fontana, A., & Frey, J. H. (1994). Interviewing: The art of science. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 361-376). Thousand Oaks, CA: Sage.

Fontana, A., & Frey, J. H. (2000). The interview: From structured questions to negotiated text. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 645-672). Thousand Oaks, CA: Sage.

Frey, J. H., & Fontana, A. (1991). The group interview in social research. *The Social Science Journal*, *28*, 175-187.

聚组数据(Grouped Data)

数据来自对某个单位或某些单位进行测量而得到的变量【Variable】。在社会科学中,单位通常是个人。其他科学中使用的单位可能是诸如猪、汽车或者诸如此类的元素。也可能有一些更大的(larger)单位,如一个县或一个国家。这类单位通常是对个体单位数据进行聚合而得到的。我们聚合个体的变量值,由此得到了更大单位的聚组数据(grouped data)。分级数据由某些变量的数据组成,如一个城市中的平均收入,或者一个选区中某一候选人得到的票数。

但是,总体数据(data on aggregates)并不总是聚组数据。如果我们将一个国家的政府类型作为一个变量,取值为民主(democracy)、专制(dictatorship)、其他(other),那么数据明显是有关这个国家整体的。但是任何国家的取值是组成这个国家的人民的特征,而不是对这个国家中的个人变量取值进行聚合得到的。

聚组数据存在于各种各样的聚合层次上。可以是一个普查统计区域的聚组数据,还可以是一个国家的聚组数据,还可以是一个州的聚组数据。这里重要的是这样一个事实,对不同层面进行聚合会从数据分析中得到不同的结论。不同层面的数据会得到大小不同的相关系数【Correlation】,相关系数取决于数据是在县的层面上聚合而成的(此时数据是关于全国 3 000 多个县的),还是在州的层面上聚合而成的(此时我们有的是全部 50 个州的数据)。因此,从某一个聚合层面上得到的分析结论只适用于这个层面。多数情况下,它们不适用于更低或更高的聚合单位。

尤其是,从聚合数据分析中得到的结论不一定适用于个体层面。从聚组数据中得到的结论被误认为适用于个体层面,这被称为区位谬误【Ecological Fallacy】。在社会学中,鲁宾逊(Robinson, 1950)在其开创性的文章中澄清了这一点。他用数学方式显示了从聚组数据中得到的区位相关系数和从个体数据中同样的两个变量的相关系数可能在大小和方向上都完全不同。这一现象也被称作"辛普森悖论(Simpson's

paradox)”。

在只存在聚组数据，个体数据没有被观察到时，情况会变得更糟。在投票中，我们知道各个选区中某个候选人得到的票数以及选区的其他特征，但是我们不知道每个人的投票情况。人们曾经尝试构建出从聚组数据中恢复个体层面数据的方法，但是原则上来讲，没有其他信息的情况下是不可能进行这种恢复的。

在社会科学数据分析中，聚组数据的另一种常见形式是交叉分类表（即用分析变量的类别进行交叉分类后的观察）。此类交叉分类表被称为列联表【Contingency Table】，常常用对数线性模型【Log-Linear Model】进行分析。

——Gudmund R. Iversen

（高勇译校）

*也可参见数据【Data】。

参考文献

Borgatta, E. F., & Jackson, D. J. (Eds.). (1980). *Aggregate data: Analysis and interpretation*. Beverly Hills, CA: Sage.

Iversen, G. R. (1973). Recovering individual data in the presence of group and individual effects. *American Journal of Sociology*, *79*, 420-434.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*, 351-357.

生长曲线模型(Growth Curve Model)

生长曲线模型是分析众多案例变化轨迹的一类技术。因此，它们适用于纵贯数据或定群【Panel】数据，这些数据中同一案例被反复观测。约翰·威沙特(Wishart, 1938)提供了生长曲线模型的较早应用，他拟合多项式生长曲线用以分析猪的增重。兹维·格里利斯(Griliches, 1957)查看了美国不同地区杂交玉米的生长，并对这些生长参数的预测值进行建模。这些早期工作以来，生长曲线模型在一般性上大大增加，扩展到了无数的应用领域。

生长曲线模型中的“生长”一词反映了这些程序源自生物学科，这些学科中研究的有机体一般会生长，一条生长曲线可以适用于每个有机体。随着这些技术扩展到社会科学和行为学科，“生长”一词显得越来越不合适，现在有称这些模型为“潜曲线模型”或“潜轨迹模型”的倾向。我们会使用“潜曲线模型”一词，以表明研究对象可以生长、衰退，或者可以遵循线性增长以外的其他模式。

社会科学家和行为科学家从两种方法视角来处理潜曲线模型。一种将潜曲线模型处理为多层次模型（层级模型）的特例，另一种将其处理为结构方程模型【Structural Equation Modeling, SEM】的特例。虽然这些方法路数各有不同，但从多层次模型和SEM中得到的模型和估计量有时完全相同，通常比较接近。这些方法路数之间的一个基本区分是，潜曲线模型是条件性的，还是非条件性的。下面两节介绍这两类模型。

非条件模型

非条件的潜曲线模型是指对单一结果进行重复测量，对此变量的轨迹进行建模。我们可以将模型构想为包括层次1方程和层次2方程。层次1方程为：

$$y_{it} = \alpha_i + \beta_i \lambda_t + \varepsilon_{it}$$

其中 i 表示案例或观察，t 表示数据的时间

或时段。y_{it}是第 i 个案例在 t 时点的结果变量或重复测量变量，α_i是第 i 个案例的截距，β_i是第 i 个案例的斜率，λ_t是观察时间的函数，通常假定为 $\lambda_t=t-1$。在本例中，模型假定结果变量完全可以用线性趋势来把握。非线性趋势也存在，我们将在最后一节讨论模型扩展时简要提及。最后，ε_{it}是第 i 个案例在 t 时间段的随机扰动项，其均值为 $0(E(\varepsilon_{it})=0)$，且独立于 α_i，β_i和 λ_t。本质而言，层次 1 方程假定样本中每个案例有各自的生长轨迹，其中截距（α_i）和斜率（β_i）决定了轨迹并且在案例之间互有差异。模型的这一部分与许多典型的社会科学统计模型有所不同，即每个个体被允许遵循不同的轨迹。

层次 2 方程将截距与斜率处理为"因变量"。在非条件模型中，层次 2 方程给出群体的截距均值和斜率均值，以满足以下条件：

$$\alpha_i = \mu_\alpha + \zeta_{\alpha i}$$
$$\beta_i = \mu_\beta + \zeta_{\beta i}$$

其中 μ_α是截距均值，μ_β是斜率均值，$\zeta_{\alpha i}$是扰动项，表示第 i 个案例的截距与截距均值的离差，$\zeta_{\beta i}$也是 β_i的扰动项，定义与此类似。扰动项均值为零，彼此可以相关。一般而言，研究者需要至少三个数据时段来识别所有的模型参数。

上述层次 1 和层次 2 方程在多层次模型和 SEM 模型的路数中具有相同形式。在潜曲线模型的 SEM 传统中，路径图是表示这些模型的另一种方式。图 1 就是非条件潜曲线模型的路径图，包括了层次 1 和层次 2 方程，数据包括了四个时段。在这个路径图中，圆（或椭圆）表示潜变量，观察变量在方块或矩形中，扰动项没有外框。单箭头表示箭头起点的变量影响箭头指向的变量，双箭头表示相连的变量间存在相关。路径图中值得注意的一个特征是，它把随机截距和随机斜率表示成了潜变量。它们是潜变量，因为我们无法直接观察到 α_i和 β_i的取值，但是我们可以估计其取值。在潜曲线模型的 SEM 方法路数中，时间趋势变量被处理成因子负载。

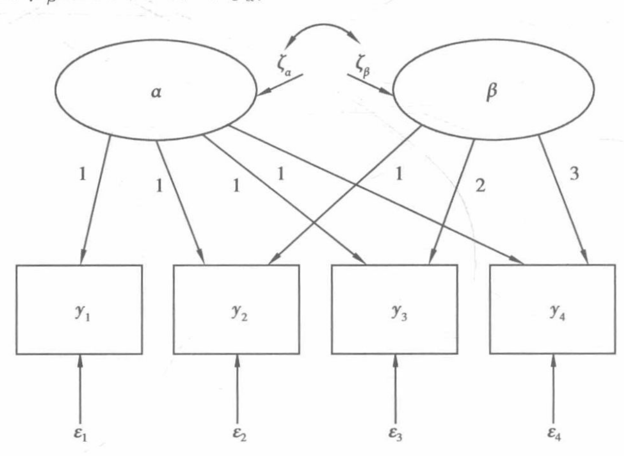


图 1　非条件的潜曲线轨迹模型

如果数据是连续的，通常会假定重复测量服从多元正态分布。此时最大似然估计量（以下简称 MLE）可用。如果观察变量不是来自多重正态分布，在 SEM 的文献中，方法学者建议对 MLE 和显著性检验进行对渐进有效的修正和调整。如果重复测量是不连续的（如二分的或定序的），就需要使用其他估计量。

为了说明非条件潜曲线模型，让我们考虑宾夕法尼亚州 616 个社区在 1995 年前 8 个月的犯罪总指数对数的双月数据。数据来自联邦调查局“犯罪统一报告”，可以用以查看该年这段时期内的犯罪轨迹。我们估计潜截距的平均值为 5.086，表明在第一个时点（一月到二月）这些社区的平均犯罪率对数为5.086。潜斜率的平均值 0.105 表明，犯罪率对数在这段时期内每 2 个月平均增加 0.105 个单位。这两个参数都高度显著（$p<0.01$）。显著的截距方差 0.506 表明年初这些社区中的犯罪水平具有相当大的变异性，而显著的斜率方差 0.007 表明在这 8 个月中犯罪率变化程度的变异性。

条件模型

潜曲线模型的第二种主要类型是条件模型。与非条件模型一样，我们可以将模型构想为包括层次 1 和层次 2 方程。除了在包括时变协变量的情况时，层次 1 方程和非条件模型中的完全一样（即 $y_{it}=\alpha_i+\beta_i\lambda_t+\varepsilon_{it}$）。差异发生在层次 2 模型中。条件潜曲线模型中的层次 2 方程不仅包括截距和斜率的均值以及与均值的离差，而且包括有助于决定截距和斜率取值的变量。例如，我们假定两个变量，x_{1i} 和 x_{2i}，通过影响案例的截距和斜率而影响到了它随时间而呈现的轨迹。

$$\alpha_i=\mu_\alpha+\gamma_{\alpha x1}x_{1i}+\gamma_{\alpha x_2}x_{2i}+\zeta_{\alpha i}$$
$$\beta_i=\mu_\beta+\gamma_{\beta x_1}x_{1i}+\gamma_{\beta x_2}x_{2i}+\zeta_{\beta i}$$

其中的 μ_α 和 μ_β 现在是它们各自方程的截距，那些 γ 是回归系数，表示在控制了方程中其他变量的条件下，变量 x 一个单位的变量对随机截距（α_i）和随机斜率（β_i）的期望影响。我们以相同方式阐释每个方程的扰动项，并且作出相同的假定，此外还需假定扰动项与变量 x 都不相关。图 2 是作为 SEM 表示出来的条件层次 1 和层次 2 方程的路径图。

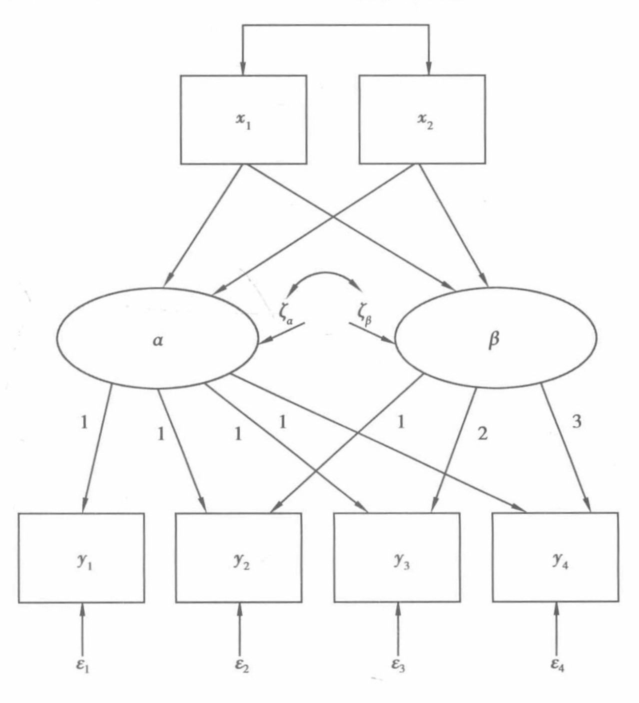


图 2　条件潜曲线轨迹模型

条件模型的优势在于,研究者可以判别哪些变量影响截距或斜率,因此更深入地理解什么影响着个体轨迹。上面条件模型中的层次 2 方程只有两个 x,但实践中研究者可以用任意数目的 x。

如果 $\zeta_{\alpha i}$,$\zeta_{\beta i}$ 和 ε_{it} 服从多元正态分布,对条件模型就可以使用 MLE,估计量取决于 x 的取值。和非条件模型中一样,如果分布假定不成立,可以对 MLE 进行调整以便进行显著性检验。

回到暴力犯罪指数上,我们将每个社区的人口密度和贫困率包括进来。贫困率对潜截距的影响估计值为 0.02 ($p<0.01$),这表明排除人口密度的因素,贫困率上一个百分点的差异导致犯罪率对数在时间始点上 0.02 个单位的差异。但是,贫困对潜斜率的影响估计值为-0.002,这表明排除人口密度的因素,贫困率上一个百分点的增加将使犯罪率对数每 2 个月的变化幅度降低 0.002。人口密度对潜截距和潜斜率都有正向影响(虽然后者刚好没有显著性),表明排除贫困率的因素,人口密度的增加将提升在时间始点上的犯罪水平,而且还会略微提升犯罪随时段呈现出来的轨迹的斜率。

扩展

非条件潜曲线模型和条件潜曲线模型有很多扩展。一种是允许结果变量有非线性的轨迹。做到这一点的方式包括拟合时间上的多项式、转换数据,或者在 SEM 路数中使得与时间趋势对应的因子负荷不受约束,以使这些负荷得到估计而非固定为常数。此外,重复测量的自回归扰动项或自回归效应也可以加入到模型中。研究者可以引入时变预测变量,视其为第二条潜曲线,或者视其为直接影响作为潜曲线模型部分的重复测量的一系列变量。直接最大似然估计或进多重插值法可以处理缺失值。研究者不仅发展了涉及有多个指标的重复潜变量的相关技术,还发展了处理非连续的(如类别的)结果变量的相关技术。此外,最近的研究已经探讨了在潜曲线模型中潜类或潜群的探测。此外有多种模型拟合测度可用,特别是在 SEM 的文献中。这些模型在多层次模型语境下的深入讨论,可以参见劳登布什和布雷克(Raudenbush & Bryk, 2002);从 SEM 视角进行的讨论,可以参见博伦和柯伦(Bollen & Curran)(in press)。目前可用于估计潜成长曲线的软件包,对于 SEM 路数来说有 AMOS, EQS, LISREL 和 Mplus,对于层次建模路数来说有 HLM, MLn 和 SAS 中的 Proc Mixed。

——Kenneth A. Bollen
Sharon L. Christ
John R. Hipp
(高勇译校)

参考文献

Armitage, P., & Colton, T. (Eds.). (1998). *Encyclopedia of biostatistics*. New York: John Wiley.

Bollen, K. A., & Curran, P. J. (in press). *Latent curve models: A structural equation approach*. New York: John Wiley.

Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica*, *25*, 501-522.

Kotz, S., & Johnson, N. (Eds.). (1983). *Encyclopedia of statistical sciences*. New York: John Wiley.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Wishart, J. (1938). Growth-rate determinations in nutritional studies with the Bacon pig, and their analysis. *Biometrika*, *30*, 16-28.

哥特曼度量(Guttman Scaling)

社会科学中的许多现象都不能用单个题项或变量直接测量。但是,研究者必须开发兼具效度与信度的对这些理论构念【Construct】的测量工具,来努力测量这些要研究的现象。度量【Scaling】就是研究者将多个题项或变量组合起来以反映感兴趣的现象的过程。社会科学中使用很多种度量模型【Model】。

最著名的一种就是哥特曼度量。哥特曼度量也被称为尺度分析(scalogram analysis)和累积度量,关注于一组题项是否测量了某一单个的理解构念。它达到上述目标的方式是,将题项和被试都沿一个潜在的累积维度根据其强度进行排序。例子有助于澄清哥特曼度量的独特之处。假定有10份国防部的拨款提案在国会进行投票——这些提案的差异只在于拨给国防支出的钱数从100 000 000~1 000 000 000美元不等,每份提案之间相差1亿美元。如果只需知道每位国会议员支持的提案数量就能够预测每位国会议员对每份提案如何投票,那么这些提案就构成了一个(完美的)哥特曼量表。例如,量表得分为8,这意味着这位国会议员支持从100 000 000~800 000 000 美元的提案,但不支持从900 000 000~1 000 000 000 美元的提案。与此类似,得分为2意味着这位议员只支持100 000 000~200 000 000 美元的提案,而反对其他8份提案。在此意义上,哥特曼度量是将题项(在本例中就是10份拨款提案)和被试(本例中就是国会议员)都沿一个潜在的累积维度根据其强度(在本例中就是给国防部的拨款钱数)进行排序。

很少有完美的哥特曼量表;事实上,哥特曼度量预计到了完美模型或理想模型会被违反。因此,问题成了经验数据在多大程度上与完美的哥特曼模型相违背。两种主要的方法可用于决定对于完美模型的背离:(a)哥特曼在1944年提出的误差最小化;(b)基于爱德华在1948年著作提出的偏离完全可复制方法。根据误差最小化标准,误差量等于:设若把观测到的回答都变换成某个理想的回答模式,需要把观测到的正向回答变成反向回答的最小数目,或需要把观测到的反向回答变成正向回答的最小数目。偏离完全可复制方法所计得的误差数量更多,因此根据尺度图理论(Scalogram theory),它是更精确地描述数据的方法。基于此理由,偏离完全可复制方法优于误差最小化方法。

——Edward G. Carmines
James Woods
(高勇译校)

参考文献

Carmines, E. G., & McIver, J. P. (1981). *Unidimensional scaling* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-024). Beverly Hills, CA: Sage.

Edwards, A. (1948). On Guttman's scale analysis. *Educational and Psychological Measurement*, *8*, 313-318.

Guttman, L. L. (1944). A basis for scaling qualitative data. *American Sociological Review*, *9*, 139-150.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

H

光环效应(Halo Effect)

光环效应也被称为“外貌魅力偏见(physical attractiveness stereotype)”或者“美丽即优秀原则(what is beautiful is good principle)”,在最具体的层面上,指的是人们对更具外貌魅力者的人格特质评价更高的习惯性倾向。光环效应也常在更为广义的层面上被使用,用来描述一种受人喜爱的个性或者一些令人渴望的具体特点所具有的广泛影响,由此对于这个人在任何维度上的判断都产生了偏差。这样,当我们赞许他人时,往往是感觉胜过了认知。

由某人的某一方面引发评判偏差的原理是基于核心特质理论:阿施(Asch, 1946)发现,一个被介绍为“热情”的人比被介绍为“冷淡”的人更可能在人格特质的许多方面被给予正面评价。

戴恩,伯奇德和沃尔斯特(Dion, Berscheid, & Walster, 1972)要求学生们去评价照片中人的个人品质。比起那些相貌不端或相貌平平者,相貌良好者在其人格特质、职业地位、婚姻竞争力、社会或职业幸福方面都被给予了更高评价。这种影响不因评判者或被评判者的性别而有所改变。

反向的光环效应也可能发生,即拥有令人尊敬的特质者也被认为更有魅力。负面的光环效应指的是由某人的相貌不端而引起的负面偏差。光环效应在诸多情景中都稳定存在,无论其文化、年龄或人的类型。它影响社会能力和智力能力的程度比影响道德正直或者调适能力的程度更甚(Eagly, Ashmore, Makhijani, & Longo, 1991)。

对个体的工作绩效进行的评估,往往会被光环效应所影响而变得无效:同事给出的评判通常反映的是被评估者的魅力或者令人喜爱程度,而不是他或她的实际绩效的客观测量。光环效应也会扭曲对于被告人罪行及其应受惩罚的判断。可以对评估者进行特定培训来减少光环效应,也可以采用特别设定的问题,或者包括一些控制题项来估计这种偏差。

在实验室情景中,美丽程度可以作为一种真正的自变量【Independent Variable】来操控,可以随机地(如通过照片)分派给每个人,在此情景中,有外貌魅力的人拥有更好的个人品质这样一种刻板印象通常是错误的。但是,在现实生活中,对于某些特质来说,它可能是对的。例如,美丽者可能事实上拥有更多的自信和社会交往技能,由此产生了自我实现预言(self-fulfilling prophecy)。

——Lionel G. Standing

(高勇译校)

参考文献

Asch, S. E. (1946). Forming impressions of personality.

Journal of Abnormal and Social Psychology, *41*, 258-290.

Dion, K. K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*, 285-290.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good but… A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*, 109-128.

霍桑效应(Hawthorne Effect)

霍桑效应指的是在行为田野实验【Field Experiment】中由于研究对象意识到自己正在被研究而带来的方法上的人为效应(参见研究过程中的人为现象【Artifact in Research Process】)。这种意识导致他们对于数据收集过程这一社会条件作出反应,而不是对于研究者试图研究的实验处理作出反应。类似于实验室研究中的"豚鼠效应",这种事实扭曲降低了实验的内在效度【Internal Validity】。霍桑效应这一术语也可以被用于管理背景中,指工人由于受到监督者特别关注而绩效提高。有些研究者将这两种用法混淆,错误地将由于霍桑效应而导致的绩效提高解释为他们实验希望达到的结果。

这两种效应都是1920年代和1930年代在西部电器公司的霍桑工厂进行的一系列研究中观察到的(Roethlisberger & Dickson, 1939)。研究者试图调查各种状况对工作绩效的影响,却惊奇地发现每种新变量都提高了生产率水平,更令人惊奇的是,在所有工作状况的改善都消除之后,生产率仍然维持在较高水平。研究者得出结论,这种意外后果是由进行对照实验这一尝试引发的附带变化引起的。

虽然对于事实扭曲的确切来源和原初实验中对此是否有证据支持,一直存在争论,但是霍桑研究首先提出资料表明了这种社会事实扭曲会干扰研究结论的可能性。霍桑效应已经被广泛认为是一种在实际场景中进行的田野实验的潜在干扰,尤其是在教育、管理、护理和其他医学场景中的社会研究中。许多社会科学学科的研究者都会不断试图控制霍桑效应,或者明确说明它可能会导致的对其研究结论的误解。

除了防止其发生的程序,研究中事实扭曲的程度通常可以用包括对照组【Control Group】的方法来解决,然后测量不进行处理的对照组和"霍桑组"之间在因变量【Dependent Variable】绩效上的差异。用来操控产生霍桑对照组的条件就是那些通常被认为产生霍桑效应的变量之一;诸如特别的关注、一项新奇却无意义的任务、引导参与者意识到自己是实验对象等。教育研究中与这些霍桑控制程序相关的效应规模【Effect Size】已经通过元分析(参见元分析【Meta-Analysis】)被发现是很小且不显著的(Adair, Sharpe, & Huynh, 1989),表明这些对照组在评估事实扭曲方面并不是很有效。

虽然缺乏证据支持对其控制有效,在确定霍桑效应的确切本质方面也存在困难,但是田野实验中事实扭曲的可能性仍旧存在。霍桑研究的重要意义在于表明研究结果是多么容易在不经意间被实验中的社会因素所扭曲。在现代研究中,霍桑效应这一术语常常在这一宽泛的意义上使用,来指涉那些源自参与者对于实验本身进行反应的非特定的事实扭曲效应。

——John G. Adair

（高勇译校）

参考文献

Adair, J. G., Sharpe, D., & Huynh, C. L. (1989). Hawthorne control procedures in educational experiments: A reconsideration of their use and effectiveness. *Review of Educational Research*, *59*, 215-228.

Jones, S. R. (1992). Was there a Hawthorne effect? *American Journal of Sociology*, *98*, 451-468.

Roethlisberger, F. J., & Dickson, W. J. (1939). *Management and the worker*. Cambridge, MA: Harvard University Press.

风险率(Hazard Rate)

风险率，又称风险函数（hazard function）或风险比（hazard ratio），是主要在生存分析【Survival Analysis】和事件史【Event History Analysis】中出现的一个概念。本词条关注它在这些文献中的用法和估计。但是，在赫克曼（Heckman, 1979）提出的用来修正样本选择性偏倚【Selection Bias】的多元【Multivariate】建模方法中，风险比的估计量也发挥了关键性作用。

风险率的数学定义始于如下假定：一个取值范围在 $t_{min} \leqslant T \leqslant t_{max}$ 中的连续随机变量【Random Varible】T，其累积分布函数（简称 CDF）为 $F(t)$，相应的概率密度函数为 $f(t) \equiv dF(t)/dt$。随机变量 T 常常被认为是对时间的测量。累积分布函数的补集被称为生存函数（survivor function），其定义为 $S(t) \equiv 1-F(t)$。它给出了 $T \geqslant t$ 的概率（即从 t_{min} 生存到了 t 的概率），所以它也被称为生存概率。风险率（或风险比）$h(t)$ 的定义如下：

$$h(t) \equiv \frac{f(t)}{S(t)}$$

粗略而言，风险率说明了，在给定 $T \geqslant t$ 时，当 Δt 值收缩为 0 时，随机变量 T 落入一个无穷小区间 $t \leqslant T \leqslant t+\Delta t$ 的概率除以 Δt 的值（数学术语来讲，就是当 Δt 趋近于 0 时取极限）。

概率密度函数和生存函数都不可能为负，所以风险率也不可能为负。风险率常常被认为与概率类似，当 T 是一个离散随机变量时，有些作者甚至称其为“离散（时间）风险率”（而不是一个概率）。但是，对于连续随机变量 T，风险率可能大于 1，此外不同于概率，它不是一个与单位无关的数字。它的量度是 T 的量度的倒数。例如，在事件史分析中，T 可能代表的是一个人经历某些事件（结婚、死亡）的年龄，或者是某人进入或脱离某种状态的持续时间（开始或结束一份工作）。在这些情况下，风险率就可以用“每年”或“每月”这样一些单位来表示。

生存函数在 t_{min} 时等于 1，因此风险率在 t_{min} 时等于概率密度函数。此外，生存函数是单调非增函数且在 t_{max} 时等于 0（除非 T 的概率分布是“不全”的），因此当 t 增大时风险率是概率密度函数更大的倍数。

在人口学中，生命表【Life Table】或者对在某一具体离散区间 $[u,v)$ 的风险率的保险精算估计【Estimation】已经使用了很长时间，现在仍然运用在许多实践应用领域中。其定义为：

$$\widehat{h(t)} = \frac{d(u,v)}{(v-u)\left\{n(u) - \frac{1}{2}\left[d(u,v) + c(u,v)\right]\right\}}$$

$$u \leqslant t < v$$

其中 $d(u,v)$ 是 t 落入区间 $[u,v)$ 的观察数目;$n(u)$ 是到 u 时仍然存活的观察数(即 $T \geqslant u$ 的观察数);$c(u,v)$ 是在区间 $[u,v)$ 中截删的观察数(对截删观察的定义与讨论,请参见截删和删节【Censoring and Truncation】)。等式右端的分母中有一个1/2,这是因为区间 $[u,v)$ 中的观察被假定为在区间中均匀分布。

另一种广泛使用的估计量源自对风险率积分的 Nelson-Aalen 估计量:

$$\widehat{h(t)} = \frac{d(u,v)}{(v-u)n(u)}, u \leqslant t < v$$

考克斯和奥克斯(Cox and Oakes, 1984)非常清晰而简明地介绍了上述风险率估计量以及其他常见的非参数风险率估计量。

风险率的多元模型可以用多种方式来分类。有些是完全参数模型;例如,风险率可以被假定为解释变量【Explanatory Variable】和随机变量 t 的特定函数,函数形式可以表达为对 t 的简单 Weibull 模型(Weibull model)或 Gompertz 模型(Gompertz model)。部分参数模型(partially parametric model)被广泛应用,特别是由考克斯提出的比例风险率模型(Cox, 1972)。在这些模型中,风险率可以表达为冗余参数 $q(t)$ 和解释变量的参数函数 $\exp[\beta' x(t)]$ 的乘积。

完全参数模型之外,不仅有比例风险率模型,而且还有非比例风险率模型。对于上述风险率模型及其他风险率模型的讨论,以及相应的社会科学示例,请参见 Tuma & Hannan, 1984, Chaps. 3-8。完全参数模型通常用最大似然法【Maximum Likelihood】来估计;考克斯的比例模型通常用部分似然法来估计。在完全参数模型和部分参数模型中,解释变量既可以是不随时间而改变的(time invariant),也可以是时变性的(time varying)。后一种类型要么需要解释变量在各个时点上更为复杂的数据,要么在没有解释变量在各个时点上的完整观察数据时,需要对解释变量的变化方式进行相当严格的假定。

对数据的探索性分析通常会表明,非比例风险率模型比比例风险率模型对经验数据的拟合更好。由吴和图曼(Wu & Tuma, 1990)提出的局部风险率模型(local hazard rate models)提供了在样本量较大时的非比例性建模方法。与之关系密切的一个概念是转移率【Transition Rate】,但有关转移率的文章指的是另一种方法。

——Nancy Brandon Tuma

(高勇译校)

参考文献

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B, 34*, 187-220.

Cox, D. R., & Oakes, D. (1984). *Analysis of survival data.* London and New York: Chapman and Hall.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica, 47*, 153-161.

Tuma, N. B., & Hannan, M. T. (1984). *Social dynamics: Models and methods.* Orlando, FL: Academic Press.

Wu, L. L., & Tuma, N. B. (1990). Local hazard models. In C. C. Clogg (Ed.), *Sociological methodology 1990* (pp. 141-180). Oxford, UK: Basil Blackwell.

诠释学(Hermeneutics)

诠释学的字面意思是"使隐晦的转换为易懂的",但它翻译成英文后,意义往往成了"阐释(to interpret)"或"理解(to understand)"。作为一种诠释圣经的方法,诠释学诞生于17世纪的德国。之后,它被应用于其他文本中,特别是那些隐晦、陌生、符号性的文本。目标是发现隐藏于文本之中的意义与意图,包括那些甚至连作者也没有意识到的意义与意图[被称为语言学(philological)诠释学]。在社会研究中,诠释学关心的是对有意义的人类行动进行阐释。

现代诠释学的起源

弗里德里希·施莱尔马赫(Friedrich Schleiermacher,1768—1843)为现代诠释学奠定了基础。正是由于他将诠释学视为理解语言中任何话语的科学,诠释学的关注重点才从分析过去文本转移到处于某种文化或某一历史时期的人如何掌握处于另一种文化或另一历史时期的人的体验的问题(被称为普遍阐释学)。

对于施莱尔马赫来说,理解具有两个维度——语法维度和心理维度。语法诠释指的是理解语言本身。心理诠释力图重建产生文本的行动,需要把自身置于作者的头脑中以便知晓作者写作时的考虑和意图。施莱尔马赫认为置身事外的诠释者比作者更能够掌握和描述"总体性(totality)"。

虽然诠释学的产生背景是对圣经和其他文本的诠释,但它逐渐被视为理解人类文化生活和物质生活中所有重大表述的核心原则。这一转变的发起者是威廉·狄尔泰(Wilhelm Dilthey,1833—1911),他认为对人类行动的研究应当基于理解方法(verstehen)以便掌握参与者的主观意识,而对自然现象的研究应当寻求因果解释(erklären)。他认为自然科学的方法并不适用于人文科学(human sciences),提出了在人文科学中如何取得客观性【Objectivity】的问题。他致力于表明人文科学中采用的方法、路数、类型也可以确保客观性和有效性【Validity】。他对这一问题提供的答案是否令人满意,则尚存争论。

在狄尔泰的早期著作中,他希望人文科学的基础应当建立在描述性心理学之上,即对于意识进行经验描述而毫不考虑因果解释。他相信心理学能够为社会科学提供基础,如同数学是自然科学的基础一样。包括文化在内的所有人类产物,都被认为源自精神生活。但是,狄尔泰逐渐意识到了这种立场的局限性,从关注个体的精神生活转向了基于意义的社会生成系统进行理解。他开始强调社会背景和他称之为"客观精神"(objective mind,即人类精神的客观化或外在化)的作用,或者在历史中沉积下来的"由精神生成的世界"(如今社会科学家称之为文化)的作用。

狄尔泰认为现象必须置于其更大的整体中,正是从整体中现象获得了其意义;部分一方面从整体中获取意义,另一方面部分又赋予了整体意义。从其外在化的产物中发现理所当然的意义,然后再从意义出发理解这些产物,这种双重过程被称为诠释学循环(hermeneutic circle)。诠释者努力重建文本产生的社会状况和历史情境,然后置文本于这些状况和情境之中。尽管方法是诠释学循环,但狄尔泰仍然坚持客观理解是人文科学最终目标的看法。

狄尔泰坚持认为理解人的基础在于生命本身,而不是理性推测或者形而上学理

论。生命或者人类体验,为我们提供了进行这种理解所必需的概念和类型。他认为人类体验中最基本的形式是活的体验(lived experience),即直接的、原生的、非反思性的体验。在狄尔泰看来,一个人或一个职业观察者理解人类产物的能力是基于如下信念的:所有的人都有共通之处。但是,他也承认一个群体的表达可能在另一个群体的成员看来根本无法理解;他们是如此迥异以至于不能互相理解。另一方面,他们可能是如此熟悉以至于根本不需要解释。

马丁·海德格尔(Martin Heidegger, 1889—1976)受到了狄尔泰的影响,奠定了现代诠释学另一分支的基础。和狄尔泰一样,他也想建立一种能够从生命本身出发揭示生命的方法。海德格尔著作的核心思想是,理解是一种存在方式而不是一种求知方式,理解是一个本体论【Ontology】问题而不是一个认识论【Epistemology】问题。它关心的不是我们如何建立知识,而是人类如何在世界上存在。对于海德格尔来说,理解是嵌在社会关系的结构当中的,解释只是把这种理解用语言明确表达出来而已。因此,理解是所有人都能够做到的事情。

海德格尔认为,对历史的理解不可能存在于历史之外。认为不言自明确实存在,其实只是没有认识到这种所谓的不言自明所赖以发生的前提条件,误将其认为是理所当然的了。所有的理解都是暂时的,任何人都不可能跨出历史或者其社会世界之外。

再总结一下,早期的诠释学兴起是为了解决对文本缺乏理解的问题,目的是要发现文本的含义。施莱尔马赫将强调的重点从文本转移到了理解某种文化或某个历史时期的人如何掌握另一种文化或另一个历史时期的人的体验。他提出了心理诠释的方法,即需要去再次体验文本作者或者讲话者的精神过程。这需要使用诠释学循环,即从碎片化的片断中去掌握未知的整体,然后再用整体来理解每个片断。接下来,狄尔泰将强调的重点转到了为人文科学建立一种普遍性的方法上,一种和自然科学方法同样严格和客观的方法。他从心理诠释转到了意义的社会生成系统,从内省性的心理学转到了社会性反思上,从精神过程的重建转到了外在化的文化产物的解释。对于施莱尔马赫和狄尔泰来说,诠释者的偏见会不可避免地扭曲其理解,因此必须把自身从社会历史背景中抽离出来。海德格尔不同意这种看法,认为理解是人类存在的基础,因此也是普通人的任务。他认为理解不可能脱离历史而存在,人不可能跨出其生活的社会世界或者历史背景。

两个分支

在客观诠释是否可能这个问题上,这些早期学者的观点分歧表明了两种分化的观点。一种观点基于施莱尔马赫和狄尔泰的著作,通过诠释学寻求诠释的一般方法原则。它致力于对历史和社会生活形成客观的诠释,这些历史和社会生活是超越和外在于人类存在的。另一种观点基于海德格尔的著作,将诠释学视为对于所有理解方式的本质和必要条件的哲学探索,认为客观有效的诠释是不可能的。海德格尔认为理解是人类的日常存在的内在部分,因此是普通人而非专家的任务。虽然这些传统都一直存在,但是后者更占优势,后来又被汉斯-格奥尔格·伽达默尔(Hans-Georg Gadamer)进一步发展了。

伽达默尔(Gadamer, 1989)对于社会科学中诠释学方法的进一步发展并不感兴趣,对于文本诠释也并不特别关心。他关注的是所有理解方式中所共有的东西,他提出了三个问题:“理解”如何可能?理解能够产生什么样的知识?这种知识的地位是什么?

对于伽达默尔来说,理解的关键在于掌

握“历史传统”，在于从特定时间和特定地点（如文本写作的时间地点）来理解和看待世界。它需要采用一种允许文本自己说话的态度，同时也要认识到必须从其他来源中去“发现”文本置身其中的传统。

从人类学和社会学的学科背景来看，历史传统可以理解为“文化”或“世界观”，而文本可以理解为社会参与者之间或者社会参与者和研究者之间的对话记录。伽达默尔的立场要求我们不只看到说了什么，而且要看到说的时候哪些东西被视为理所当然，要看到语言的日常含义，看到对话发生的情境的日常含义。目标在于听到“弦外之音（to hear beyond the mere words）”。

伽达默尔方法的另一个重要特点是认识到理解其他传统或文化产物的过程并不能够脱离诠释者置身其中的文化。他对狄尔泰对人类产物进行“客观”诠释的努力持有批评态度。对于伽达默尔来说，文本或者历史行动都只能从诠释者的意义视野中得到诠释，当诠释者的意义视野与这种行动或文体的意义视野融合时，视野就被拓宽了。理解的过程包括了这样一种“视野融合”，即诠释者自己的意义视野在与既有视野的诠释对话中通过问答的辩证过程而被改变了。诠释者与文本进行对话，这一过程既改变了文本，也改变了诠释者。诠释者不带偏见、思想开放地接近文本，目标并不是去发现文本的真正含义；他（她）不是一个认识者，而是一个其他传统向他敞开的体验者。

对于伽达默尔来说，诠释学沟通了我们熟知的世界与陌生的世界中存在的意义之间的鸿沟。与其他视野的碰撞能够使诠释者意识到自身根深蒂固的假定、偏见、意义视野，而诠释者在此之前可能对此并没有意识到；理所当然的假定被加以批判性的自我审视，真正的理解因此成为可能。

伽达默尔认为理解其他人的话并不是要“进入到”他们的头脑中，再次经历他们的体验。因为语言是理解的普遍媒介，理解就是语言之间的翻译。但是，每一种翻译同时也是一种诠释。每一次对话都要假定两位发言者使用同样的语言，能够理解对方所说。但是诠释性对话通常发生在不同语言之间，从我们通常认为的“外语”，到同一种语言由于时代变迁导致的变化，或同一种语言在不同方言土语之间的差异。诠释学的任务不是正确地掌握另一种语言，而是在不同语言之间进行调解。作为“视野融合”的理解是通过语言发生的，语言使得过去与现在的调解、渗透、转化成为可能。不论是两个人之间的对话，还是诠释者和文本之间的对话，共同语言必须被创造出来。同海德格尔一样，伽达默尔感兴趣的也是本体论而非认识论，感兴趣的是确立所有理解方式的共同之处而非方法问题。

诠释学为诠释主义（interpretivism）的一些分支奠定了基础，如马克斯·韦伯和阿尔弗雷德·舒茨（Alfred Schütz）的著作。它也为外展逻辑（the logic of abduction）奠定了基础。

——Norman Blaikie

（高勇译校）

参考文献

Bauman, Z. (1978). *Hermeneutics and social science.* London: Hutchinson.

Blaikie, N. (1993). *Approaches to social enquiry.* Cambridge, UK: Polity.

Gadamer, H-G. (1989). *Truth and method* (rev. 2nd ed.). New York: Crossroad.

Outhwaite, W. (1975). *Understanding social life: The method called verstehen.* London: Allen & Unwin.

Palmer, R. E. (1969). *Hermeneutics: Interpretation theory of Schleiermacher, Dilthey, Heidegger, and Gadamer.* Evanston, IL: Northwestern University Press.

异质性(Heterogeneity)

统计学研究的就是异质性,所以这是个很大的题目。统计学模型通常会区分可观察的异质性和不可观察的异质性,用预测变量或解释变量来掌握可观察的异质性来源,不可观察的异质性来源则与误差项联系起来,误差项通常假定是独立于可观察协变量的。异质性也源于个体观察服从不同的模型,因此有了个体独特模型(subject-specific model)和总体平均模型(population-average model)的区分。由于这些理念最早流行于人口学应用中,我们从生存数据或者时间-事件数据的角度来描述这些理念。在这个方面的开创性论文是由沃佩尔、曼顿和斯托拉德(Vaupel, Manton, and Stallard, 1979)写出的。

试考虑一个同质性总体,其中一件诸如死亡这样的事件,如果还没有发生,那么它在时间(或年龄)t 发生的风险为 $\lambda(t)$,这一点对个体和对总体都成立。再考虑一个异质性总体,假定对个体 i 来说风险为$\theta_i\lambda(t)$,其中的 $\theta_i>0$,它代表一个人的脆弱性,是不可观察的,其分布【Distribution】的均值为 1。在这样的异质性总体中,在年龄 t 的平均风险【Hazaro】不是 $\lambda(t)$,而是 $E(\theta_i \mid T_i>t)\lambda(t)$。这个乘数是活到年龄 t 的个体的平均脆弱性,它是小于 1 的,并且随时间 t 而降低。这表明,即使对每个个体来说风险是个常数【Constant】,总体【Population】的平均风险将会降低,因为较虚弱者在早年就已经死亡了,留下来的都是经过筛选的更为健康者。

在异质性总体中观察到的风险不同于每个个体面临的风险,这一事实就引出了有趣的悖论。分析者们因此就尝试在模型中除了引入可观察的协变量外,再引入脆弱性的影响来控制不可观察的异质性。估计【Estimation】通常需要设定风险的参数形式和脆弱性的分布。估计值会非常依赖于对不可观察因素的分布假定【Assumption】,因此赫克曼和辛格(Heckman and Singer, 1984)提出了一种对 θ_i 的分布的非参数估计量【Estimator】。但是估计对于风险的参数形式也同样很敏感。遗憾的是,我们不可能同时放松这两个假定,因为那样模型就不可识别了。具体而言,不可观察的异质性在没有协变量的模型中是与负向的持续性依赖(negative duration dependence)混杂在一起的,在有协变量的模型中是与通常的风险成比例假定混杂在一起的;详情可以参看 Rodríguez(1994)。

如果每个个体都有重复性事件,而且脆弱性对这些事件是持续性的,或者数据是如孩子嵌套于母亲这样具有层级性的,而且脆弱性在同一家庭成员中是相同的,那么情况就会好得多。在这两种情形中,脆弱性模型是完全可识别的。对于多元生存模型的讨论,请参看 Hougaard(2000)。这些模型可以被视为多层次分析【Multilevel Analysis】模型的特例。经典的脆弱性模型对应于随机截距模型,其中响应变量在各组中的均值水平有所不同。也可以考虑随机斜率模型,其中协变量对于结果的影响(即处理效应)在各组中有所不同。或者,脆弱性独立于可观察协变量的假定也可以放松,这需要将 θ_i 处理为固定效应而非随机效应(参看固定效应模型【Fixed Effects Model】)。

——Germán Rodríguez

(高勇译校)

参考文献

Heckman, J. J., & Singer, B. (1984). A method for minimizing the impact of distributional assumptions in

econometric models for duration data. *Econometrica*, *52*, 271-319.

Hougaard, P. (2000). *Analysis of multivariate survival data*. New York: Springer-Verlag.

Rodríguez, G. (1994). Statistical issues in the analysis of reproductive histories using hazard models. In K. L. Campbell & J. W. Wood (Eds.), *Human reproductive ecology: Interaction of environment, fertility and behavior* (pp. 266-279). New York: New York Academy of Sciences.

Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, *16*, 439-454.

异方差性(Heteroscedasticity)

参见异方差性【Heteroskedasticity】。

异方差性(Heteroskedasticity)

要证明最小二乘【Ordinary Least Squares,OLS】是最佳线性无偏估计量【Best Linear Unbiased Estimator,BLUE】,一个必要的假定是误差的方差对于所有观察都是恒定的,即同方差性【Homoskedasticity】。如果误差项并不恒定,这一假定被违背时,就产生了异方差性问题。这种情况发生时,OLS参数估计仍是无偏的,但是估计参数的方差不再有效。一个经典的例子就是收入对储蓄的回归(图1提供了一个假想的例子)。我们能想到,储蓄会随着收入提高而增加。这个回归是异方差性的,因为低收入者没有什么钱来储蓄,都只有相对很小的储蓄量,误差项也就很小(他们都位于回归线附近)。虽然,高收入者有更多钱可以储蓄,但是有人选择将多余的钱投入其他用途(例如休闲),所以他们的储蓄水平差异很大,误差项就很大。

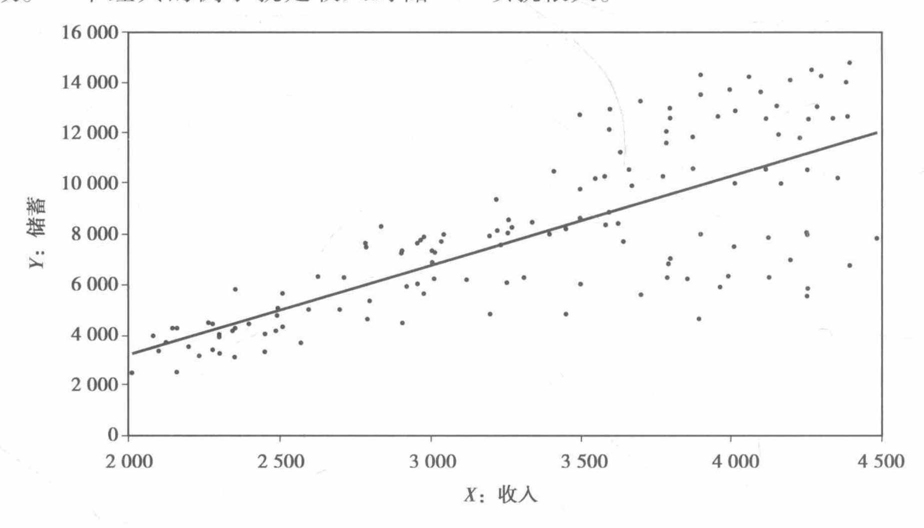


表1 异方差性下的回归

有很多种方法用以探测异方差性。最简单的方法是把残差$\hat{u}_i = (y_i - \hat{y}_i)$和每个自变量的关系画在图上进行观察。如果残差的绝对值在所有自变量的不同取值中都保持恒定,那么异方差性就不是问题。如果在回归线周围的误差项的散布在某一自变量上存在变异(如图2所示,这显示了我们假想例子中的情况),就可能存在异方差性。在任何情况下,都应该进一步进行异方差性的正规检验。

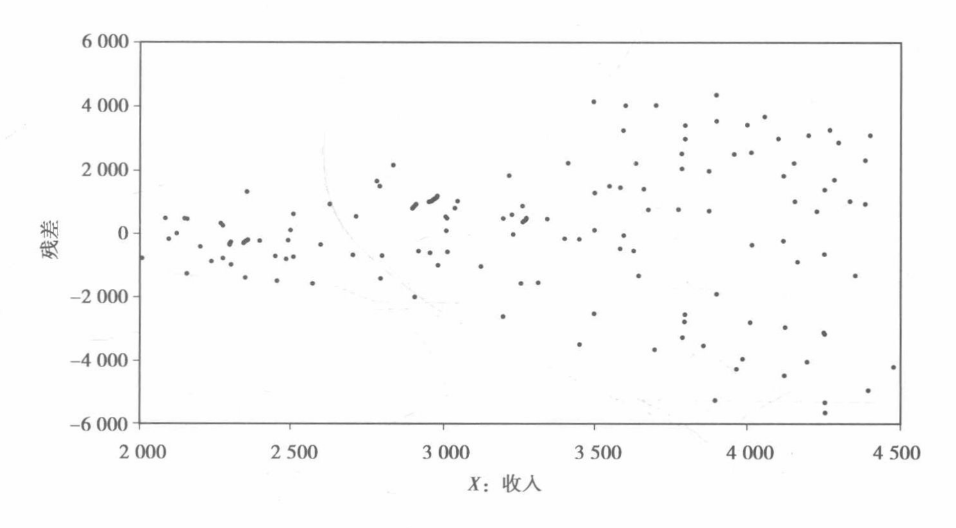


表2 残差可视化

可以使用的正规检验包括 Goldfeld-Quandt 检验和 Breusch-Pagan 检验。前者会根据引起异方差性的变量的大小对每个观察进行排序,一些处于中间的观察被略去,然后对剩余的两组观察分别进行回归。它们的残差平方和之比,被用来检验异方差性是否存在。后者更具普遍性,因为它并不要求了解引起异方差性的变量。

如果检验显示异方差性存在,接下来的第一步是再去细究残差,以便发现能否表明遗漏了变量。那些估值过度的个案可能都有一个共同特质,这个共同特质可以在模型中加以控制。在本例中,测量个体对于储蓄(或者休闲)的内在价值重视程度,可能会部分解决问题。高收入-低储蓄者可能比一般人更不看重储蓄(或者更看重休闲)。如果不能对模型设定进行这些明显的改动,就有必要转换变量,进行广义最小二乘【Generalized Least Squares,GLS】了。用来纠正异方差性的 GLS 程序会对观察在违反假定的自变量上进行加权,权数为其误差项的倒数,这也常被称为加权最小二乘法【Weighted Least Squares,WLS】。如果进行了转换,并且其他回归假定没有被违反,转换后的估计量是最佳线性无偏估计量。

——Kevin J. Sweeney

(高勇译校)

参考文献

Berry, W. D. (1993). *Understanding regression assumptions*. Newbury Park, CA: Sage.

Gujarati, D.N. (1995). *Basic econometrics* (3rd ed.). New York: McGraw-Hill.

Kennedy, P. (1992). *A guide to econometrics* (3rd ed.). Cambridge: MIT Press.

便捷方法(Heuristic)

用来使某一概念或某种方法更便于理解的工具、手段或准则。例如,为了帮助学生学习显著性检验【Significance Testing】,可以给他们提供如下便捷方法:一般而言,t 比值的绝对值大于等于 2.0 的系数在 0.05 水平下会是显著的。在整理许多次回归【Regression】运行生成的 t 比值时,这样的便捷方法就会很有用。

——Michael S. Lewis-Beck

(高勇译校)

启迪调查法(Heuristic Inquiry)

启迪调查法是现象学调查的一种方式,它将个人体验与研究者的洞察力放在重要的位置上。对于感兴趣的现象,调查者要问:"我对这种现象的体验是什么? 那些也强烈体验到这种现象的其他人对这种现象的本质体验是什么?"

人本主义心理学家克拉克·穆斯塔卡(Moustakas, 1990)提出了启迪调查法,他用这个词来涵盖他认为在调查人类体验中最重要的过程。"启迪(heuristic)"来源于希腊词 heuriskein,意思是"发现"或"找到"。它隐含了一种内在找寻的过程,通过这种找寻人们发现自身体验的本质与意义,提出进一步调查和分析的方法与程序。"通过这一过程,研究者的自我得以展现,同时越来越深入地理解了现象,研究者也体验到不断增长的自我觉知和自我了解。启迪过程包括了创造性的自我提升(self-processes)和自我发现"(Moustakas,1990,p.9)。

在现象学的更大框架中,存在着启迪调查法的两个核心要素:(a)研究者对于要研究的现象必须有个人体验,必须要有强烈兴趣;(b)对这一现象也有强烈体验的其他人(合作研究者)要参与到探究中来。道哥拉斯和穆斯塔卡(Douglass and Moustakas, 1985)强调说,"启迪关切的是意义,不是测量;是本质,不是表象;是性质,不是数量;是体验,不是行为"(p.42)。

启迪调查法的特别贡献在于,它在很大程度上证明研究者的个人体验、反思、洞察力在研究中是合理正当的,把这些因素放在重要的位置上。在研究者对所研究现象也有强烈的体验和反思时,他们通过与合作研究者(coresearchers)分享反思与调查,从而逐渐理解现象的本质。研究者和研究参与者在共同致力于澄清一种重要人类体验的本质、意义、实质的过程中,培育出了一种心灵相通之感。

启迪调查法的严谨性来自对自我和他人的系统观察和不断对话,以及对合作研究者的深度访谈。

启迪调查法源于现象学【Phenomenology】,但在如下四个主要方面与现象学有所不同:

(1)启迪调查法强调心灵相通与发展关系,而现象学在分析体验时更加鼓励超脱。

(2)启迪调查法通过描绘个人意义来寻求本质意义,而现象学强调对体验的结构进行确切的描述。

(3)启迪调查法包括研究者的直觉和只可意会的理解,而现象学要提取出体验的结构。

(4)"在启迪调查法中,研究参与者是可见的……现象学则以体验本质告终;启迪

保留了体验中人的本质”(Douglass & Moustakas, 1985, p. 43)。

启迪调查过程要经过沉浸、孵化、阐明、澄清和创造性综合的系统步骤,最终显现体验的本质(Moustakas, 1990)。

——Michael Quinn Patton

(高勇译校)

参考文献

Douglass, B., & Moustakas, C. (1985). Heuristic inquiry: The internal search to know. *Journal of Humanistic Psychology*, *25*, 39-55.

Moustakas, C. (1990). *Heuristic research: Design, methodology, and applications*. Newbury Park, CA: Sage.

Moustakas, C. (1994). *Phenomenological research methods*. Thousand Oaks, CA: Sage.

Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage.

层级(非)线性模型(Hierarchical [Non] Linear Model)

层级线性模型(HLM)是多层次分析【Multilevel Analysis】的主要程序的基础。它是一般线性模型【General Linear Model】在具有层次嵌套结构的数据中的扩展,在此种数据中每个层次水平都是未解释变异的来源之一。一般线性模型可以用公式表示为:

$$Y_i = b_0 + b_1 X_{1i} + \cdots + b_p X_{pi} + E_i$$

其中 i 代表个案;Y 是因变量;X_1 到 X_p 是自变量(或解释变量);E 是未解释部分,通常称为残差或者误差项;b_0 称为截距项;b_1 到 b_p 是回归系数。有时常数项 b_0 可以从模型中省略掉。标准假定:残差 E_i 是独立的、服从正态分布的随机变量,期望值为 0,方差相同;数值 X_{hi} 是固定且已知的量。

举例来讲,嵌套结构可以是学生嵌套于班级(两水平结构),学生嵌套于班级再嵌套于学校(三水平结构),学生嵌套于班级再嵌套于学校再嵌套于国家(四水平结构),等等。这样的嵌套结构中的元素被称为单元(如此例中的学生就是第一水平单元,班级是第二水平单元,以此类推)。对于两水平嵌套结构来说,通常用双标来标注,例如 Y_{ij} 表示在第 j 个班级中第 i 个学生的因变量。两水平嵌套结构的层次线性模型可以表示为:

$$Y_{ij} = b_0 + b_1 X_{1ij} + \cdots + b_p X_{pij} + U_{0j} + U_{1j} Z_{1ij} + \cdots + U_{qj} Z_{qij} + E_{ij}$$

第一部分,$b_0 + b_1 X_{1ij} + \cdots + b_p X_{pij}$ 称为固定部分,其结构和解释与一般线性模型的第一部分完全相同;解释变量的双标只是标注方式不同而已。被称为随机部分的第二部分则有所不同。第二水平单元间的未解释变异用 $U_{0j} + U_{1j} Z_{1ij} + \cdots + U_{qj} Z_{qij}$ 来表示,其中 Z_1 到 Z_q 是解释变量(通常是 X_1 到 X_p 的一个子集,但不是一定如此);U_{0j} 被称为随机截距;U_{1j} 到 U_{qj} 被称为随机斜率(也可以没有随机斜率,即 $q=0$)。变量 U_{0j} 到 U_{qj} 是第二水平的残差,在不同的第二水平单元间变动,E_{ij} 是第一水平的残差。标准假定是不同水平的残差之间彼此独立,同一水平中不同单元的残差彼此独立,所有残差都服从正态分布。也可以把模型扩展为在第一水平中除了包括残差 E_{ij} 外,还包括随机斜率;这是表示第一水平残差中存在异方差性【Heteroskedasticity】的一种正规方式(参见

Snijders & Bosker, 1999, Charp. 8)。

模型可以扩展到嵌套水平多于两层的情形。对于三层嵌套水平,不同水平的单元分别用 i,j,k 来表示,公式如下:

$$Y_{ijk} = b_0 + b_1 X_{1ijk} + \cdots + b_p X_{pijk} + V_{0k} + V_{1k} W_{1ijk} + \cdots + V_{rk} W_{rijk} + U_{0jk} + U_{1jk} Z_{1ijk} + \cdots + U_{qjk} Z_{qijk} + E_{ijk}$$

现在的 W_{hijk} 也是解释变量,随机斜率 V_{hk} 在不同的第三水平单元 k 中变动。模型的其余部分与两水平模型相似。层次线性模型也可以扩展到未解释变异的来源并非嵌套而是交叉因素(crossed factors)的情形中(参见本书中多水平分析【Multilevel Analysis】和混合效应模型【Mixed Effects Model】的词条)。

层级非线性模型

层级非线性模型,也称为层级广义线性模型或者广义线性混合模型,是对广义线性模型【Generalized Linear Model】的类比扩展。它可被界定为在广义线性模型中的线性预测值中加入了高水平单元的随机残差。故而,前述对于 Y_{ij} 和 Y_{ijk} 给出的公式,在去掉了第一水平残差 E_{ij} 或者 E_{ijk} 之后,就取代了广义线性模型词条中定义为线性预测值 $\eta = X\beta$ 的那些向量元素。

——Tom A. B. Snijders

(高勇译校)

参考文献

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Hodder Arnold.

Raudenbush, S. W., & Bryk, A. S. (2001). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

可信度层级(Hierarchy of Credibility)

霍华德·贝克尔(Becker, 1967)提出的这个术语,用来指社会中不同群体在其观念被认为是合乎情理的能力上有所差异的现象。这个术语指涉的是,社会的许多领域中都有清晰层级的倾向:警察/罪犯,经理/工人,教师/学生。有研究表明,当社会科学家采用上述对子中地位较低的群体的世界观时,他们更容易被指责为持有偏差。产生上述现象的原因即在于可信度层级,那些处于优势地位者借此被社会大众认定为有权力和资源来定义事物存在的方式。

——Alan Bryman

(高勇译校)

参考文献

Becker, H. S. (1967). Whose side are we on? *Social Problems*, *14*, 239-247.

高阶(Higher-Order)

参见阶【Order】。

直方图(Histogram)

直方图、条形图【Bar Graph】、饼图【Pie Graph】以简单、可视的方式展示结果。条形图用于类别(定类)变量,图中的每个条形展示了一些变量的取值。直方图执行的功能与此相同,但是用于连续变量【Continuous Variable】(定序、定距或比例的分布)的分组数据。条形之间紧紧相连,以此反映背后分布的连续性质。直方图或条形图用于展示变量(在不同时段或不同地段间的)趋势时,要优于饼图。饼图更适于显示总体的百分比构成。

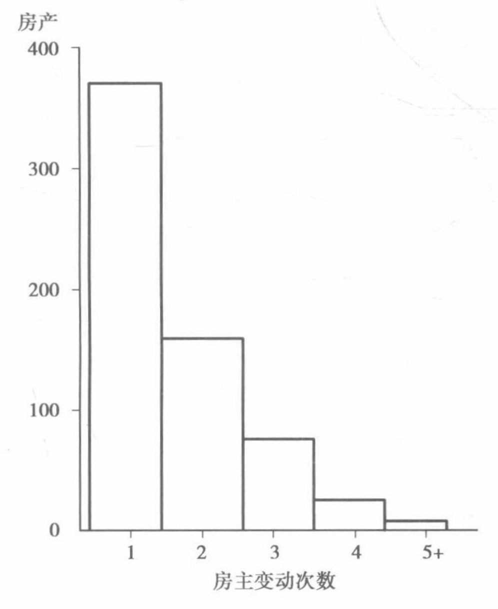


图 1　2001 年米德尔斯堡(Middlebrough)小镇中房主变动次数

图 1 展示了过去一年中那些已经有过一次房主变动的房产中记录下来的迁移(即房主变动)次数,这是英国东北部的一个小镇中心,它在主导产业衰落以后已经走下坡路了,但现在又有所复兴。这个图显示出,接近 150 座房产在一年间经历了二次转手,有相当数量房产的所有权甚至变化更加频繁(此图中的数据来自 Hobson-West & Sapsford, 2001)。除了在描述数据上很有用,这种图也有助于确定变量的整体分布形态,如与正态分布的偏离程度,在本例中这种偏离是非常明显的。

——Roger Sapsford

(高勇译校)

参考文献

Hobson-West, P., & Sapsford, R. (2001). *The Middlesbrough Town Centre Study: Final report*. Middlesbrough, UK: University of Teesside, School of Social Sciences.

历史方法(Historical Method)

历史方法指使用一手的历史资料来回答问题。由于资料的性质取决于要回答的问题,资料可以包括人口记录,如出生证明和死亡证明;报纸文章;信件和日记;政府记录;或者甚至可以包括建筑图纸。

历史资料的使用提出了几个更大的问题:

(1)对于提出的理论问题来说,资料是否适合?

(2)这些资料当初是如何收集的,或者在收集时何种意义嵌入这些资料当中?

(3)如何诠释这些资料,或者现在这些资料保有何种意义?

理论问题

社会科学家对当代议题提问的一种方式就是比较当代社会与以往社会的各个方面。如赖特·米尔斯所言,要理解今天的社会生活,就必须提出以下问题:"我生活在何种社会当中?""在人类社会的行进中,这个社会位于何处?"这些问题必然隐含着理论,隐含着指导研究者在何处发现答案的一系列相关命题。事实上,大体来说,任何历史问题都隐含着理论(any historical question implies a theory),因为没有理论,社会科学家就是在毫无章法地收集一堆毫无结构的事实。

我强调明确的理论是历史方法的核心,以此来彰显历史学家使用的历史方法和社会科学家使用的历史方法之间的关键区别。虽然历史学家也使用理论,但他们的理论假定并不总是得到明确的发展。他们的理论重点常常被包含在他们所处的专业分类中——如"军事历史学家""外交历史学家""社会历史学家"——因为这些分类隐含地表明某一特定群体的活动(以及某一种特定权力或者丧失权力)是这些时代的分析核心。在历史专业内部,如同在所有学术专业内部一样,任何一个范畴的核心地位都会随时代而改变。在20世纪后期,社会史的进展(包括如女性史和非洲裔美国人史这样的专史的进展)就曾经宣称,理解那些历史上处于无权地位的人,甚至完全被人忽视的人,是掌握我们时代本质的关键。

适宜的资料

此类范畴重点的转变也意味着新的资料形式。例如,新闻报道对理解农民、工厂工人、奴隶、女仆的生活的用处,就比它对写作一位军官的传记的用处更小——虽然新闻故事也可以告诉我们那些有更大权力的人们是怎样看待工厂工人或奴隶的。与此类似,通过比较大众媒体和非洲裔美国人媒体,我们就有希望能够理解欧洲裔美国人和非洲裔美国人对城市暴乱的不同反应。但是除非此类报道被置于一个理论背景和比较背景下,研究者就只是有希望理解而已。

遗憾的是,许多着手进行历史调查的社会科学家并未接受历史学的训练。这一缺乏部分地意味着,社会科学家在阅读历史文献时可能既不知道这些文献所争辩的具体问题,也不知道历史学家要讲的隐含理论。没有这些背景知识,就很难知道该记下什么,或者甚至很难知道在随处可见的笔记卡片或计算机文档中写下什么。与此类似,不理解生活在特定时间地点的特定群体中的人们的生活,就不可能理解某一特定资料的意义或重要性。

没有接受过历史学训练也提出了另一个问题:一个社会科学家可能不知道与一个研究问题相关的档案位于何处。参加历史编纂学的课程,咨询学识渊博的历史学家,查找《历史摘要》(*Historical Abstracts*),在网络上搜索,都是了解其他人已经使用到哪些资料和可以得到哪些类型的资料的办法。

示例

一些社会科学家用归纳法而不是演绎法工作。如果一位社会科学家采用归纳方式,他或她必须了解背景知识。举一个看似无关紧要的例子——构建一个家谱(一件不牵涉理论的事情)。为了画出家谱,我回想我父母、叔叔、姑姑的童年故事。然后,我参观我家的墓园,手里拿着笔和纸。我之所以这么做,是因为我知道在每个人的英文姓名下面,我会发现其希伯来文的名字,那个名字会遵循以下格式"首名、中间名,首名、中间名的儿子或女儿",如"Chava Perl bat [daughter] of Yisroel Avram"。通过确定Yisroel Avram的墓碑,我就可以知道他父亲

的名字,以此类推。有一块墓碑没有给出死者父亲的名字,而是用意第绪语(错误的语言)写着"一位好人的女儿"。一位长辈亲戚曾经告诉我一个有关墓碑是意第绪语的女人的家族秘密:她嫁给了她的叔叔。在查看了墓地的位置之后(我知道其风俗),我能够推断出她叔叔的姓名,也知道了她母亲的姓名。如果不掌握这些背景信息,我就不能构建出这个家谱。这个家谱就是等待着理论框架的资料。

其他背景问题

在其他方面,背景问题同样也很困难。阅读历史,需要了解历史学家在论辩什么问题(如果历史学家没有自己的观点,他们也就根本不会去写作了)。其他一些非虚构作品也有其他问题:阅读 19 世纪的女性日记需要提出以下问题,这种技能是在学校习得的吗?教师是否会将其作为作业的一部分而读到这些日记?在这些文字记录的写作中涉及什么样的自我审查(或外部审查)?背景——什么被说出来和什么不被说出来——决定着历史资料中的意义,正如在当代社会的民族志研究中一样。背景引导着研究者发现与理论相关的资料,而理论指引着研究者理解相关的背景,这一点对于任何定性研究都是适用的。

解读

对于解读资料来说,理论尤其重要,因为所有的理论都内含着对资料的本质存在的认识论【Epistemology】立场。宽泛而言,至少有两种不同的解读方式。我称其为回溯推理(reproduction)与表述(representation)。我用"回溯推理"指如下观点,认为历史叙事准确地捕捉到了某一特定时间和地位的本质。我在此包括了经验主义认识论,如强调"历史的行进"的那些认识论。"表述"指各种各样的后现代【Postmodernism】观点,认为历史写作和社会科学写作都是"文本",和他们分析的资料一样,都是在权力斗争中或有意或无意地进行站边的政治文档。但是,对于这些思想派别间的区别进行讨论,已经偏离了我们对历史方法的讨论。

——Gaye Tuchman

(高勇译校)

参考文献

Himmelfarb, G. (1987). *The new history and the old*. Cambridge, MA: Harvard University Press.

Marcus, G., & Fischer, M. (1986). *Anthropology as cultural critique*. Chicago: University of Chicago Press.

Scott, J. W. (1988). *Gender and the politics of history*. New York: Columbia University Press.

Tuchman, G. (1994). Historical social science: Methodologies, methods, and meanings. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 306-323). Thousand Oaks, CA: Sage.

使其不变(Holding Constant)

参见控制【Control】。

同方差性(Homoscedasticity)

参见同方差性【Homoskedasticity】。

同方差性(Homoskedasticity)

要满足一般最小二乘【Ordinary Least Squares, OLS】假定,误差【Error】必须具有稳定方差。这样均匀的方差被称为同方差性。如果不能满足这一假定,即如果误差不具有均匀方差,就出现了相反的异方差性【Heteroskedasticity】条件,此时 OLS 不再是模型参数【Parameter】的最佳线性无偏估计量【Best Linear Unbiased Estimator, BLUE】。具体而言,如果误差是异方差的,估计出的模型参数仍然是无偏的【Unbiased】,但它们将不再是方差最小的估计量。

要理解均匀误差方差这一概念,试考虑两个变量 X 与 Y 之间的关系,以及在一个回归中 Y 对 X 的条件均值与方差。如果因变量【Dependent Variable】Y 与自变量【Independent Variable】X 相关,那么 Y 的均值将会因 X 的取值不同而有变化。这就是 X 给定时 Y 的条件均值,写作($Y|X$)。众所周知,Y 对 X 的回归线直接穿过这些条件均值。我们可以考虑 Y 在不同条件均值时的变异。这被称为条件变异,用 var[$Y|X$]代表。这些条件变异也被称为条件方差函数(skedastic function)。在同方差性时,条件变异不会因 X 的取值变化而变化,它是均匀的。这种均匀的变异用 σ^2 表示(如图 1 所示)。在 X 的取值给定时 Y 的条件均值($Y|X$)会随 X 增大而增加。但是围绕这些条件均值的误差散布却会保持恒定。

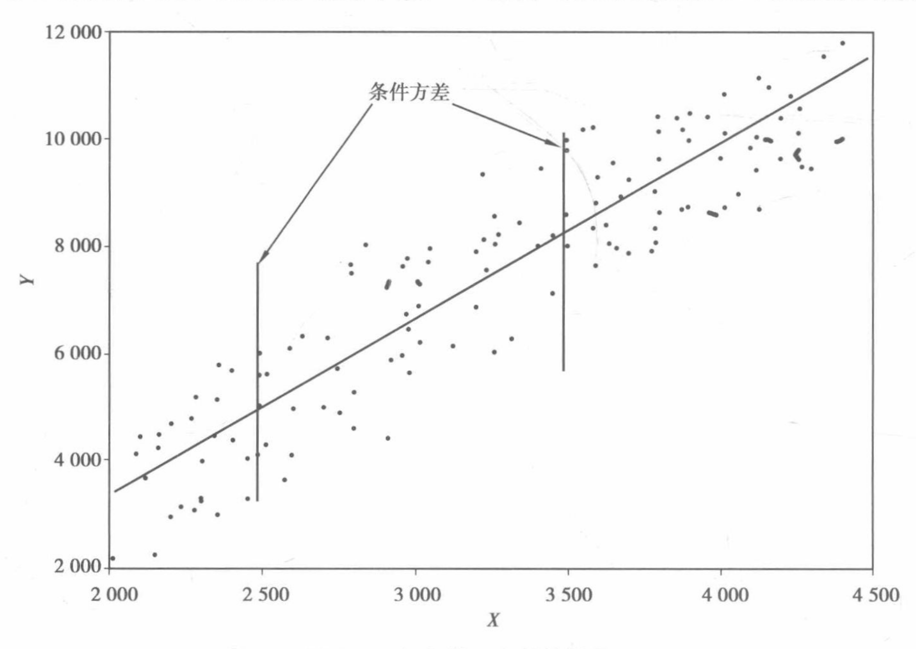


图 1　回归中的同方差性误差

如果误差是同方差性的,而且彼此之间不存在自相关【Autocorrelation】,就被称为是"球形的(spherical)"。发果扰动项【Disturbance Term】是球形的,误差方差-协方差矩阵的期望值【Expected Value】E(**UU**′)就可以简化为 $\sigma^2\mathbf{I}$,就可以很容易地得到模型参数及其标准误【Standard Errors】的 OLS 公式(参看球形假定【Sphericity Assumption】)。如果扰动项不是球形的,那么 OLS 就有问题了,此时需要用到广义最小二乘【Generalized Least Squares, GLS】解。

误差就是因变量 Y 的预测值和观察值的差值,因此它的单位和 Y 的单位是一样的。但是,没有理由指望 Y 的整体变异 var(Y)会等于 var($Y|X$),即使条件变异是均匀的 σ^2。可以预期,我们的模型会解释掉 Y 的整体变异的一部分,因此要满足 OLS 假定,只需要残差的条件变异是均匀的。

——Brian M. Pollins

(高勇译校)

参考文献

Greene, W. H. (2003). *Econometric analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Gujarati, D. N. (2003). *Basic econometrics* (4th ed.). Boston: McGraw-Hill.

Kennedy, P. (1998). *A guide to econometrics* (4th ed.). Cambridge: MIT Press.

人本主义和人本主义研究(Humanism and Humanistic Research)

人本主义研究就是认为研究核心应当是人、人的意义、人的行动的研究。它通常带有很强的道德框架,强调尊重人,致力于在全世界范围内提升人类状况。它更多地认同自己是人文学科而非科学,科学往往成为它批评的对象。

各种不同形式的人本主义有着悠久的历史。希腊的智者派(Sophists)和苏格拉底把哲学从天上带到了地上(如西塞罗所言)。文艺复兴运动前的早期人本主义者,如伊拉斯谟,虽然并未发现他们的人本主义和宗教不能相容,但确实意识到了人在创造和控制世界中起到了重要的作用。意大利文艺复兴中,对人(而非上帝)的创造的研究变得越来越重要。法国哲学家,如伏尔泰,赋予了人本主义坚实的世俗基础。艾尔弗雷德·麦克伦格·李(Alfred McLung Lee, 1978)是一位人本主义社会学的拥护者,他将人本主义置于一系列广泛的宗教、政治、学术运动中(p.44),把它和诸多领域联系起来,如共产主义、民主、平等主义、民粹主义、自然主义【Naturalism】、实证主义【Positivism】、实用主义、相对主义、科学,以及包括古代异教、印度教、佛教、犹太教、罗马天主教、新教、伊斯兰教等各式各样的超自然主义(pp.44-45)。确实每个人对它都有各自不同的理解。

在社会科学中存在明确人本主义派别的学科有心理学(Abraham Maslow, Jerome Bruner, and Robert Coles)、人类学(Gregory Bateson, Margaret Mead, Robert Redfield)、社会学(C. Wright Mills, Peter Berger,以及符号互动论者)。例如,人本主义心理学批评多数心理学研究忽视人性、琐屑无关,实验虽然精致,却对当下整个人类面临的重大道德危机以及处理方式视而不见。在社会学中,人们常常把人本主义研究和考察人们生活中意识建构的研究联系起来。

与威廉·詹姆斯的《宗教经验之种种》(*Varieties of Religious Experience*)或者玛格丽

特・米德（Margaret Mead）的《三个原始部落的性别与气质》（*Sex and Temperament in Three Primitive Societies*）一样，人本主义社会科学通常会承认意义的多样性。它常常使用的方法包括实地研究、生活故事、定性研究，使用的材料包括文学、传记、摄影、摄像，它努力将生活如其所是地熟稔于心。人本主义研究倾向于避免抽象，而采用一种更为实用的、讲求实际的路数。虽然它通常对人类理性抱有很强信心，但是它也承认情感和感觉的作用，并且它极其强调人们在创造更好世界和更好生活的过程中作出的伦理选择。

人本主义研究受到来自多个方面的批评。科学家会嘲笑它缺乏客观性、确实性和技术能力。尼采和克尔凯郭尔等虚无主义者则不同意它对人类过于浪漫的看法，批评其伦理系统。许多宗教批评它，因为它并不承认人的卑微，不承认人只能依靠神的恩宠得到拯救。最近许多文化多元论者和后现代主义者也批评它，因为它对人是什么持有一种特定的西方自由主义观点。尽管有这么多批评，但它仍然是对社会科学主导的科学倾向持有反对意见的一种传统。

——Ken Plummer

（高勇译校）

＊也可参见自传【Autobiography】、诠释性传记【Interpretative Biography】、生活史访谈【Life History Interview】、生活史方法【Life History Method】、口述史【Oral History】、定性研究【Qualitative Research】、反身性【Reflexivity】。

参考文献

Lee, A. M. (1978). *Sociology for whom*. New York: Oxford University Press.

Radest, H. B. (1996). *Humanism with a human face: Intimacy and the enlightenment*. London: Praeger.

Rengger, J. (1996). *Retreat from the modern: Humanism, postmodernism and the flight from modernist culture*. Exeter, UK: Bowerdean.

人文系数（Humanistic Coefficient）

不同于从“事实”和“事物”的角度来看待世界的社会科学，社会科学中还存在另一种悠久的传统，它关心的是有意义的、存在主体间性的社会行动和价值（meaningful, intersubjective social actions and values）。马克斯・韦伯对理解【Verstehen】的强调是此类路数中最有名的一种。阿尔弗雷德・舒茨的现象学【Phenomenology】著作是另一种路数。

在1920年代和1930年代，另一位清晰阐释这种立场的著名人物是弗洛里安・兹纳涅茨基（虽然今天他已经基本上被遗忘了）。在《身处欧美的波兰农民》（*The Polish Peasant in Europe and America*）（与W. I. 托马斯合著）的方法注释和《社会学方法》（*The Method of Sociology*）中，他表达了对新康德主义区分自然系统和文化系统的赞同。自然系统是自然客观赋予的，有着独立的存在。它们独立于人的经验和活动。文化系统内在地和人的自觉经验关系密切，这些经验又是人这一主体在与他人互动中产生的。

兹纳涅茨基（Znaniecki, 1934）认为，“自然系统是客观地赋予科学家的，仿佛它们绝对独立于人们的经验和活动而存在”（p.136）。这类系统包括地球的地理构成、化学化合物、磁场等一类事物。它们可以不依赖人类意识的任何参与而存在。如他所言，“它们是由那些与人类活动无关的力量

集合在一起的”(p.136)。与此截然相反,文化系统处理的是人的自觉体验和积极的历史主体。那些体验到文化系统的历史主体确信其是真实存在的,但这种真实存在和自然系统的真实存在不是同一意义的。研究的对象总是与某人的意义相关联。兹纳涅茨基将文化资料的这种本质特征称为人文系数,“因为这些资料作为学生进行理论反思的对象,早已属于他人的积极经验了,因为正是这些积极的经验创造了这些资料”(p.136)。因此,如果忽略了文化系统,未将其视为一种人文系数,如果像研究自然系统一样去研究文化生活,那么研究者只能发现“一堆互不相连的自然事物和过程,与其开始调查的现实毫不相像”(p.137)。它是研究人类生活和文化生活的关键。

因此,社会科学的一个主要传统就是不断地强调研究这种“人文系数”的重要性,强调明白历史社会生活的参与者们建构和理解自身世界的方式的重要性,即他们的“情境定义”,他们的“一阶建构(first level constructs)”。托马斯和兹纳尼茨基的经典著作《身处欧美的波兰农民》(*The Polish Peasant in Europe and American*)就是这一传统的典范,他们利用了一整套主观材料——信件、生活故事、个案记录等——来理解移民生活经验的意义。这一经典研究预示了社会科学中个人文件、生活史、参与观察的重要发展。这些都是帮助我们理解人文系数的核心方法。

——Ken Plummer

(高勇译校)

*也可参见理解【Verstehen】、可感概念【Sensitizing Concept】、生活史方法【Life History Method】、参与观察【Participant Observation】。

参考文献

Bryun, T. S. (1966) *The human perspective in sociology: The methodology of participant observation*. Englewood Cliffs, NJ: Prentice Hall.

Thomas, W. I., & Znaniecki, F. (1918-1920). *The Polish peasant in Europe and America* (Vols. 1 & 2). New York: Dover.

Znaniecki, F. (1934). *The method of sociology*. New York: Farrar and Rinehart.

Znaniecki, F. (1969). *On humanistic sociology: Selected papers* (R. Bierstadt, Ed.). Chicago: University of Chicago Press.

假设(Hypothesis)

假设是科学方法的核心特征,特别是假说-演绎科学模型的核心要素(Chalmers, 1999)。这一模型的各种形式广泛应用于科学之中,包括社会科学。假设是从现有理论【Theory】中推导出来的陈述,它可以用特定的科学方法来进行检验。在化学或心理学中,这可能是实验;在社会学或政治学中,这可能是社会调查。假设可以被证实或证伪,它们的检验结果将会对理论产生影响,理论可能作出相应修正,然后又产生出新的假设。考虑如下简化的例子:经典迁移理论认为经济上的推拉因素是迁移的动机,行动者能够清醒地认识到这些因素。从这种理论出发,可以假设移民将会从经济停滞地区迁移到经济繁荣地区。如果经过检验,发现某些经济停滞地区反而有着较高的移入水平,那么原先的理论就必须进行修正,要么修正其经济因素的命题,要么修正行动者对此有完备知识的命题,或者两者都需要修正。

但是,假设很少能够被证明为完全正确

(证实)或完全错误(证伪),源出的理论能够被完全证实或证伪的情况就更为少见。20世纪后半叶中,对于假设以及背后的理论在多大程度上能够被证实或证伪,存在大量的争论(时至今日仍然存在)。卡尔·波普尔(Popper, 1959)坚信,从逻辑上来说,只有证伪方法才能解决问题,因为无论已经记录了多少次证实的情形,你也不能确定未来不会有被证伪的情形。但是,只要有一次证伪情形就能够说明在理论设定及其推导出的假设中存在错误。因此,在这种观念看来,科学家应当努力去推翻假设。

证伪在逻辑上是正确的,它暗含着的怀疑主义也能够使假设检验更为严格,但是特别是在社会科学中,假设不可能是全称命题(universal statement),而往往是概率性的。在上面的例子中,没有人会认为在某种特定情形下所有人都会迁移,而是认为人们会更可能或更不可能迁移。事实上,在许多情形中,一个假设是否应当认为被证实还是被证伪,往往只是取决于调查发现中态度或行动几个百分点的差异。

假设可以在不同的层面上设定。研究假设是一种语言陈述,其证实或证伪也只能用语言来陈述,如"没有""全部""部分""大部分"等。在定性研究中,假设可能被表达为一种有某种特性的社会情境,这可能通过观察来加以证实或证伪。但是在调查或实验研究中,假设检验可以得到一个发现的统计显著性,因此可以判断这一发现是由于偶然概率发生还是存在真实效应的证据。虚无假设被加以表述,如移民和拥有房产之间没有关系,但是如果事实上发现了显著关系,那么虚无假设就被拒斥,备择假设被证实。备择假设可以是一个更广的研究假设,或者其一部分(Newton & Rudestam, 1999, pp. 63-65)。

——Malcolm Williams

(高勇译校)

参考文献

Chalmers, A. (1999). *What is this thing called science?* (3rd ed.). Buckingham, UK: Open University Press.

Newton, R., & Rudestam, K. (1999). *Your statistical consultant: Answers to your data analysis questions*. Thousand Oaks, CA: Sage.

Popper, K. (1959). *The logic of scientific discovery*. London: Routledge.

假设检验(Hypothesis Testing)

参见显著性检验【Significance Testing】。

假说-演绎法(Hypothetico-Deductive Method)

通常将这种方法称为"假设方法"。它包括获取或发展出一个理论【Theory】,从中逻辑性地推导出一个假设(演绎),这为与某一研究主题相关的一个"为什么"的研究问题【Research Question】提供了一个可能答案,然后再去将它和从这一问题被调查的场景中获取的合适的资料进行比较来检验这个假设。

假说-演绎法的基础是由英国数学家、神学家威廉·希韦尔(Whewell, 1847)奠定的,他借此来反对更早的由弗兰西斯·培根倡导的归纳【Induction】方法。希韦尔与其同时代人约翰·斯图尔特·密尔争论归纳的概念。他批评密尔认为的科学知识就是从许多特殊观察中进行普遍化的观念,他向那种认为观察可以完全在没有前置概念的前提下进行的观点发起挑战。他拒斥那种认为对观察进行普遍化是放之四海而皆准的合适科学方法的理念,坚持认为假设必须在科学研究的早期阶段就发展出来以便解释观察到的现象。对他而言,除非研究者可以把观察到的现象围绕一些概念组织起来,这些现象本身没有什么意义。简而言之,他将解释的来源从观察转移到了科学家头脑中的构念(constructions)上,这些构念将用来解释观察到的现象。

这些概念化(conceptions)包括使用以前没有用于解释这些"事实"的概念或措辞。在开普勒的例子中,这就是"椭圆轨道(elliptical orbit)",对牛顿而言,这就是"重力(gravitate)"。希韦尔认为,这个过程中需要有"创造性天分";这需要猜测多种概念化方式,然后选择出正确者。他认为,如果假设能够完美地拟合事实,就不可能怀疑它的真实性。尽管有着这样一种自我确认(self-validation),他还是准备通过进行预测和合适的观察来进一步检验假设。正是在后一点上,希韦尔是卡尔·波普尔的证伪主义【Falsificationism】的先声。

波普尔(Popper, 1972)对于假设用来组织理念的概念并不感兴趣。他所关心的是假设或者猜想被提出来作为研究问题的试探性答案,然后再进行检验这样一种科学观。波普尔认为,尽管人们相信科学是从观察到理论的过程,但是像实证主义【Positivism】那样,认为可以没有一丁点理论,完全从观察出发的看法是荒谬的。观察永远是选择性的,是在一个参照系或者"期望视野(horizon of expectation)"中进行。波普尔认为,我们不能等待规律从观察中自动强加于我们之上,我们必须积极地去将规律强加于世界之上。我们必须在一定程度上妄自断定(jump to conclusion),虽然随后的观察如果显示结论错误,这些结论就会被放弃。

因此,波普尔认为是科学家在以理论的形式创造着规律。但是,态度必须是批判性的而不是教条性的。教条性的态度将使得科学家寻求对他们理论的验证,而批评性的态度包括情愿通过对理论的检验和拒斥来改变理念。他认为,我们永远不可能确认理论是否真实,我们所能希望的只是去除那些错误的理论。优秀的理论能够经受住严格的检验。

——Norman Blaikie

(高勇译校)

参考文献

Blaikie, N. (1993). *Approaches to social enquiry.* Cambridge, UK: Polity.

Popper, K. R. (1972). *Conjectures and refutation.* London: Routledge & Kegan Paul.

Mill, J. S. (1947). *A system of logic.* London: Longman, Green. (Originally published in 1879.)

Whewell, W. (1847). *The philosophy of the inductive sciences* (Vols. 1 & 2). London: Parker.

I

理想类型(Ideal Type)

理想类型是定性社会研究中用于资料分析的构念【Construct】。理想类型可以用来系统地对历史资料(包括生活史资料)进行比较性分析,这一思想最早由马克斯·韦伯提出。韦伯的理想类型方法已经在理想类型分析中被改动,以满足多个案研究的要求。通过使用理想类型,我们系统地分析生活史资料的每一个个案,最终得到结构性的解释。理想类型分析是诠释性社会研究的一个分支,这种社会研究处理的是沿多个个案轨迹组织起来的叙述性(文本)资料。

历史发展

马克斯·韦伯最早提出了理想类型的思想,用来解释社会学的个体文化现象。也就是说,韦伯(Weber,1904/1949)在致力于理解从社会现实的"无限的因果网络"中得到的单个个案材料时,提出了理解类型作为其解释方案(p.84)。阿尔弗雷德·舒茨是另一位在研究中使用了理想类型的理论家,他意识到理论类型不仅是微观社会理论的有效建构工具,而且社会生活本身也是在理想类型结构中组织起来的(Schütz,1932/1967,1955/1976)。

在这样的背景下,格哈特(Gerhardt,1985,1994)把理想类型引入定性研究中。他在两项主要的研究中采用了理想类型,系统地展示了这一概念的应用范围(Gerhardt,1986,1999)。这种方法已经在德国及其他国家确立起来(如 Leisering & Leibfried,1999)。

个案分析的层次

在研究工作的三个阶段中,个案是理想类型分析的焦点。第一,个案是通过资料处理过程确立的分析单位;第二,要根据其表现理想类型的相对能力来对个案进行选择;第三,个案解释是理想类型分析的最终目标。个案工作的三个层次如下:

1.个案重构(case reconstruction)。这意味着对资料中的所有个案进行重构,使它遵循分析方案依序排列。个案重构把个案按一种有序模式进行安排,以便每个个案都可以和资料集中的所有其他个案进行比较。

2.选择典型性个案(selection of paradigmatic cases)。这是对个案进行比较分析的结果(遵循系统性的个案重构)。如果出现了相似个案的聚类时,在每一个聚类中选择一个或多个能够集中体现这种模式特点的典型性个案。

3.个案解释(case explanation)。它是研究工作的最终目标,它在结构性解释之后得到(参看下一节)。对于韦伯和现代理想类型分析者来说,目标是系统地解释经验个案

的发展动态。

分析程序

通过两个层面的分析程序,研究者可以得到描述和结构性解释,两个层面中都会用到理想类型。描述用来构建概括个案动态过程的描述类型。解释用于理解个案材料的动态过程的结构模式。在描述层面和结构层面使用理想类型来理解社会生活模式,需要在分析程序中经过两个单独的阶段:

1. 描述性理想类型(descriptive ideal types)。在用分析方案重构个案时通过个案比较,相似个案就会实现聚类。从中选择一个典型性个案作为理想类型,就可以得到这些聚类的景象。理想类型个案集中体现了每个聚类中的个案动态过程。通过比较典型性个案(理想类型)和聚类中的其他个案,对每个聚类进行相同的程序,个案动态过程的丰富景象就浮现出来。这一景象展现了这一组中的模式,以及在比较基础上建立的所有组的模式。这考虑到了经验材料中的多种多样的比较,由此可以得到在理想类型的典型性个案中概括的一系列模式的全部场景。

2. 结构性理想类型(structural ideal types)。如果提出这样的问题:"为什么个案会这么进行?"社会结构就进入了我们的视野。这就需要考虑模式动态过程的结构解释。只有通过以理想类型为基础的结构解释,我们才能真正理解单个个案。这类解释决定了为什么个案遵循着一种结构模式的动态过程。为了选择出能够用来进行结构模型解释的理想类型个案,我们必须在研究项目的分析视角下,建立一套个案特征的名册,这个名册定义了最优个案的特征或构成。在此基础上选出的理想类型个案是在经验材料中发现的。理想类型个案集中体现了结构模式,后者与其个案的特征尽可能吻合。理想类型个案的深度分析揭示了结构模式中"纯粹"类型的结构动态过程。我们可以试着用韦伯的术语来说,理想类型个案代表了用于分析目标的结构模式。但是,从分析上来讲,不仅应当建立最优的个案,也应当建立最差的情况。从最优个案中得到的分析图景和从最差情况中得到的分析图景放在一起,才能够得到模式的丰富的全部场景。这些模式定义了在分析个案材料中包括的潜在个案动态模式的结构背景。高地位对低地位群体,多次结婚对从未结婚家庭背景,在此只提这两项内容,这些就构成了其动态过程可以使用理想类型分析来研究的社会结构模式。

理想类型个案考虑到了联结个案和结构模式的解释。因为社会结构在资料中被具体化为个案中的一个变量值,它们相对于理想类型个案的典型性也有助于理解他们的结构性发展动态过程及个体发展动态过程。比较性分析将理想类型个案和资料中的其他个案并置,这样就可以进行子群体的分析,因此这又进一步丰富了结构模式解释的发现。

示例

我们对于冠状动脉搭桥手术病人的传记研究调查了 60 个个案(使用了 240 次访谈)。在手术前选择这些个案的根据是,每个个案都满足四个预测其手术后能够重新工作的已知标准(既有社会标准,也有医学标准)。略超过一半的个案最终在手术后能够重新工作;其他个案自己选择或者最终被迫提前退休。个案重构产生了四种模式,即血管重建成功且重新工作,血管重建失败且重新工作,血管重建成功且提前退休,血管重建失败且提前退休。我们使用描述性理想类型(即四种模式中每种的典型性个案)对这四组进行进一步调查。为了解释个案

的轨迹,我们使用最佳效果的四个标准对所有个案进行了再编码(手术后没有症状,收入提高,手术后在重新工作还是退休的决策上和配偶没有矛盾,满意的医患关系)。只有两个个案全部满足四个标准,一个是重返工作的个案,一个是提前退休的个案。他们在许多解释参数上都是不同的,如社会阶层、进行手术时的年龄、婚姻状态、对健康状况的主观感知,以及在其叙述中表达出来的主要生活兴趣。那个出身上层、年龄较大、重返工作的个案,其主要生活兴趣是偏向工作和职业的;而那个出身下层、年龄较轻、提前退休的个案,其主要生活兴趣与工作毫无关系(偏向休闲和没有工作压力的老年生活)。使用跨个案比较分析的方法(比较不同参数背景下的生活轨迹,如社会阶层),这两个理想类型个案被用来解释其他个案中的事件发展。研究的主要结论之一如下:在搭桥手术后重返工作还是提前退休,主要取决于主要生活兴趣的类型,而不是血管重建的结果。

理想类型和测量

马克斯·韦伯(Weber,1904/1949)说明了理想类型的目标是测量。他进行了如下表述:"这些概念是我们通过客观可能性的范畴构造各种联系的构念,借助于这种范畴,我们想象力的充分性受到评判,这些想象力是受着事实指导和训练的。"(p. 93)遵循韦伯的传统,同时展示了韦伯的理想类型方法的理念如何可以转化成为现代定性研究的一种方法,格哈特论述了在资料中其进程与理想类型相对立的那些特殊个体传记中,如何将个案解释作为一种个案发展的测量:

> 个案解释可以作为一种示例,说明了个案中如何体现出了偶然性和不可预见的风险。正是在个案与理想类型相背离之处,其独特性方才显现。看似非理性的因素,却使得其独特的动态过程变得可以理解,因此对个案作出了解释——正如韦伯所指出的那样。(p. 117)

——Uta Gerhardt

(高勇译校)

参考文献

Gerhardt, U. (1985). Erzähldaten und Hypothesenkonstruktion. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *37*, 230-256.

Gerhardt, U. (1986). *Patientenkarrieren: Eine idealtypen-analytische Studie*. Frankfurt: Suhrkamp.

Gerhardt, U. (1994). The use of Weberian ideal-type methodology in qualitative data interpretation: An outline for ideal-type analysis. *BMS Bulletin de Methodologie Sociologique* (International Sociological Association, Research Committee 33), 45, 76-126.

Gerhardt, U. (1999). *Herz und Handlungsrationalität: Eine idealtypen-analytische Studie*. Frankfurt: Suhrkamp.

Leisering, L., & Leibfried, S. (1999). *Time and poverty in Western welfare states*. New York: Cambridge University Press.

Schütz, A. (1967). *Der sinnhafte Aufbau der sozialen Welt* [Phenomenology of the social world]. Evanston, IL: Northwestern University Press. (Originally published in 1932)

Schütz, A. (1976). Equality and the meaning structure of the social world. In *Collected Papers II* (pp. 226-273). The Hague: Martinus Nijhoff. (Originally published in 1955)

Weber, M. (1949). Die "Objektivität" sozialwissenschaftlicher und sozialpolitischer Erkenntnis ["Objectivity" in social science and social policy]. In *The methodology of the social sciences: Max Weber* (pp. 50-112). New York: Free Press. (Originally published in 1904)

唯心论(Idealism)

唯心论是一种哲学观点,它是社会科学中许多认识论立场的基础。这些认识论都认为现实是取决于精神的,或者由精神调节的。唯心论的对立面是唯实论【Realism】,唯实论认为现实是独立于精神而存在的,但是唯心主义思想内部也包括了许多有差异的观点。这些观点并不完全拒斥外在现实的存在;毋宁说它们指出了我们有关外在现实的知识存在局限。

事实上,最重要的唯心论哲学家之一康德认为,只有在存在于时空世界的背景下,我们才能够将自己视为具有精神禀赋的生物(Körner, 1955, Chap.4)。重要的是,我们只有通过对现象的感知才能了解现象,而这种感知是依赖于精神的。我们甚至不能了解事情本身,康德称其为"本体(noumena)"。康德的"先验唯心论(transcendental idealism)"产生了直接或间接的巨大影响,在马克斯·韦伯的著作中这种影响可能最为重要。他将认识问题从事物自身转移到了认识工具的完善上来,特别是"理想类型"。理想类型不是社会现象的平均,也不是社会现象中最可取的,而是一种推理方式,个体用这种方式可以通过共有的理性能力推论出,如果所有行动者都完全理性地行动将会出现的模式或者纯粹概念类型。因此,韦伯的方法论中至关重要地取决于理性,而理性是精神的产物(Albrow, 1990)。

黑格尔的唯心论与康德的唯心论极其不同。康德哲学的重点是认识论,是要完善认识工具;黑格尔强调的是本体论,是要发展出我们对精神的理解。不同于康德的本体,黑格尔相信可以通过矛盾及其解决的辩证过程来认识精神(以及作为普遍范畴的精神)——吊诡的是,这种方法最常和马克思联系在一起,而马克思是一位唯物主义者和唯实论者!

唯心论社会科学的多数表述都没有明显而直接地溯源至康德或黑格尔,它们只是源于一些前提预设:社会世界是由理念构成的;只是因为我们认为社会世界存在,它才存在;在此基础上我们再生产了社会世界。再生产的工具被认为是语言,因此唯心论支持社会科学中的"语言学转向"。彼得·温奇(Winch, 1958/1990)是一位重要的创始人,至少在英语世界中是如此,他认为社会生活中的规则遵循是由语言实践构成的,因为语言是不断变化的(并相应地塑造着文化),因此要了解一种文化,就必须去了解这种文化的语言。温奇的哲学是常人方法论【Ethnomethodology】的重要组成部分,对稍微晚近的后现代主义【Postmodernist】方法也有间接影响。

——Malcolm Williams

(高勇译校)

参考文献

Albrow, M. (1990). *Max Weber's construction of social theory*. London: Macmillan.

Körner, S. (1955). *Kant*. Harmondsworth, UK: Penguin.

Winch, P. (1990). *The idea of a social science and its relation to philosophy*. London: Routledge. (Originally published in 1958)

识别问题(Identification Problem)

在社会科学中,经验研究涉及从由数据、参数化模型、统计矩条件组成的预先设定好的系统中得到一致性的(consistent)总体参数估计【Parameter Estimation】。在此背景下,识别问题是指模型参数无法得到一致性估计的任何情形。能够得到模型参数的一致性估计的系统就被称为可识别的(identified),无法得到一致性估计的系统就是无法识别的(not identified)。有时,一个系统只能在部分参数上得到一致性估计,此时识别问题就是在参数水平上。识别问题可以粗略地分为两类:由于矩条件不足(insufficient moment condition)而引起的识别问题,以及由于模型的过度参数化(overparameterization)而引起的识别问题。但是,这两类之间存在相当程度的重叠(过度参数化的模型可以被视为矩条件太少,反之亦然)。这些现象在理论上和经验上的简要例证将在下面给予介绍。

普遍最小二乘【Ordinary Least Squares, OLS】回归可以被视为对由于矩条件不足而引起的识别问题的一种解决办法。考虑如下简单的确定性(deterministic)系统:

$$y_i = \alpha + x_i\beta + \varepsilon_i, i = 1, \cdots, n \quad (1)$$

其中 ε_i 是非随机的松弛项(slack terms)。只有方程(1)是无法决定未知参数 α 和 β 的;这里本质上存在着 n 个方程和 n+2 个未知量(n 个松弛项也被视为未知量)。因此,系统的识别还缺两个方程。但是,如果松弛项是随机误差,那么如下两个 OLS 矩条件就通过 OLS 回归提供了两个用以识别 α 和 β 的一致性估计的必要方程。

$$\mathrm{E}(\varepsilon_i) = 0(\text{零均值条件}) \quad (2)$$

$$\mathrm{E}(x_i\varepsilon_i) = 0(\text{外生性条件}) \quad (3)$$

熟悉的工具变量【Instrumental Variable】技术是解决识别问题的一个典型案例。如果方程(3)中的外生性条件不成立,那么由方程(1)和方程(2)构成的系统的识别就缺一个矩条件。如果存在对 x_i 的合适工具变量 z_i,那么如下四个额外方程:

$$x_i = \gamma + z_i\delta + v_i, i = 1, \cdots, n$$

$$\mathrm{E}(v_i) = 0$$

$$\mathrm{E}(z_i v_i) = 0$$

$$\mathrm{E}(z_i\varepsilon_i) = 0$$

再加上方程(1)和方程(2),就可以用于识别参数 α,β,γ 和 δ(实际上存在 $2n$+4 个方程和 $2n$+4 个未知量,在此 ε_i 和 v_i 也被视为未知量)。注意,如果 β=0,由方程(1)和方程(2)构成的系统也是可识别的(但此时也无关紧要),所以这个识别问题也可以被简单化地视为过度参数化(overparameterization)问题之一。

识别问题中的过度参数化类别有一个重要例子,就是联立方程问题(simultaneous equations problem),这通常要用系统中的参数约束来加以解决。在包括多个模型方程的系统中,变量是联立决定的,因此通常都不能满足方程(3)中的外生性条件。例如,如果 y_i 和 x_i 是联立决定的,z_i 是外生【Exogenous】的,那么联立模型就为如下所示:

$$y_i = x_i\beta_1 + z_i\theta_1 + \varepsilon_{1i} \quad (4)$$

$$x_i = y_i\beta_2 + z_i\theta_2 + \varepsilon_{2i} \quad (5)$$

$i=1,\cdots,n$，其中 $E(\varepsilon_{1i})=E(\varepsilon_{2i})=E(\varepsilon_{1i}\varepsilon_{2i})=E(z_i\varepsilon_{1i})=E(z_i\varepsilon_{2i})=0$。

这个系统中的两个方程都无法识别(not identified)。要理解这一点，注意方程(4)中的 y_i 通过方程(5)中的 x_i 与 ε_{2i} 相关。类似地，方程(5)中的 x_i 通过方程(4)中的 y_i 与 ε_{1i} 相关。因此，整个系统不满足 OLS 的外生性条件，不能得到一致性估计。此外，一些代数运算表明，方程(4)和方程(5)在参数意义上是等价的，因此不可能有解。如果理论表明存在参数约束 $\theta_1=0$，那么上述方程就不是等价的，z_i 就可以作为 y_i 的合适工具变量。此时，方程(5)可以使用工具变量技术得到一致性估计。方程(4)仍然是无法识别的，因为包括在数据(x_i,y_i,z_i)中的所有信息都已被耗尽(关于联立方程模型中识别问题的详尽理论处理，请参看 Greene，2003；Kmenta，1997；Schmidt，1976；Wooldridge，2003)。

曼斯基(Manski，1995)介绍了一些社会科学应用研究中关于识别问题的实例。例如，曼斯基提出的在社会项目评估上的"混合问题(mixing problem)"就是一个识别问题。也就是说，在进行社会项目的受控试验时，对实验处理组样本的处理是同质的，或者对其中不同人群的处理是有差别的，结果是有差别的。因此，在社会项目实验的结果中，社会项目在不同人群中的潜在结果是不可识别的。此外，工资决定的经济模型中的"选择问题(selection problem)"也是一个识别问题。市场工资数据通常都排除了总体中那些选择不工作者的工资观察资料；因此，如果用那些选择进入劳动力市场者的工资数据来估计整个总体(工作者与不工作者)的工资系数估计，那这个问题就是不可识别的。在上述两个例子中，可以通过将先验信息引入问题中来解决识别问题(请参见 Manski，1995)。只要把这些先验信息视为额外的矩条件，前面的混合问题和选择问题就都是由于矩条件不足引起的识别问题。

——William C. Horrace

(高勇译校)

参考文献

Greene, W. H. (2003). *Econometric analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Kmenta, J. (1997). *Elements of econometrics* (2nd ed.). Ann Arbor: University of Michigan Press.

Manski, C. F. (1995). *Identification problems in the social sciences*. Cambridge, MA: Harvard University Press.

Schmidt, P. (1976). *Econometrics*. New York: Marcel Dekker.

Wooldridge, J. M. (2003). *Introductory econometrics* (2nd ed.). Mason, OH: South-Western.

特殊知识/通则知识(Idiographic/Nomothetic)

参见通则知识/特殊知识【Nomothetic/Idiographic】。

影响评估(Impact Assessment)

影响评估需要对一项干预(intervention)的长期效果进行全面评价。

对感兴趣的现象,如一个社区发展项目,评价者的问题是:从预期效果和未预期效果两方面来说,这个项目对社区有何影响(Patton, 1997)?

影响评估给评价带来了更广的、整体性的视角。评价中第一个评估层面是资源层面或者输入层面:项目或计划在多大程度上取得了实施所需要的目标资源?第二个评估层面是实施分析:项目或计划在多大程度上如计划实施?第三个分析层面关注产出:从参与水平、完成情况、最终产出来看,项目或计划在多大程度上取得了预期目标?第四个分析层面涉及后果:项目或计划的参与者的生活在多大程度上以何种方式按预期目标得到了改善?最后,评价的最终层面是影响评估:后果在多大程度上是在长时间段中可持续的或者有所增进的,在更大的范围内(如社区中)产生了什么样的"涟漪效应(ripple effect)"?

以一个教育项目为例。评价的第一个层面是输入评估,需要考察资源的充足程度——足够的建筑物、材料、胜任的员工、交通等。第二个评价层面是实施分析,需要考察预期的课程多大程度上都被教授,是否以高质量的方式进行教学。第三个评价层面是输出,如毕业率、减少辍学率、家长参与指标、(学生、家长、教师的)满意度数据等。第四个层面是后果评价,要看成就得分、学生工作的质量、对学生在情感方面和社会方面的成长的测量。最后,影响评估要看这个教育项目给学生毕业后带来的长时间效果,给项目实施所在社区带来的效果,例如社区荣誉感、犯罪率、住房拥有方面的变化,或者由于高质量的学校为社区吸引了商业项目,或者商业项目由于学校绩效太差而搬出了社区。

就本质而言,影响评估中存在着两个维度:时间维度(time dimension)和范围维度(scope dimension)。时间维度是长时段来看:除了即刻的、短期的后果,什么效果是可持续的或者逐渐增进的?后果评价(outcomes evaluation)对于干预和参与者变化之间的直接有益联系进行评估,如知识、能力、技巧方面的增长,再如态度和行为方面的变化。与此不同,影响评估(impact assessment)考察这些后果在多大程度上能够在长时段中保持和持续。例如,一个学前教育项目的目标是帮助孩子做好第一年上学的准备。后果就是这些孩子在校第一年的表现。影响则是他们在随后几年中的表现。

影响评估的第二个维度是范围。后果评价考察的是干预和后果之间的有限的、具体的、直接的联系。影响评估在更宽广的范围内考察涟漪效应、未预期后果、副作用、背景效应(contextual effects)。影响评估既包括考察正面效果,也包括考察负面效果;既包括预期后果,也包括未预期后果。例如,一个扫盲项目预期的可测量后果是原本是文盲的成年人会阅读了。影响则包括这些成年人能找到更好的工作、社会参与更多,甚至住房拥有率更高、离婚率更低。因此,影响评估关注的不仅是直接的、即刻的因果关联,而且是更长时段的、更广范围中的效果。

这种范围更广、时段更长的视角也带来了方法上的挑战。每件事情、任何事情都是可能的调查对象,因为要寻找的不仅是计划中的、可预见的效果,还包括未预期的效果。社区通常是影响评估的分析单位(Lichfield, 1996)。全面的影响评估需要多种方法和测量办法,既包括定量的【Quantitative】也包括定性的【Qualitative】,既包括演绎的也包括归纳的。当我们进行长时间段分析时,当我们考察的是超越项目干预【Intervention】和即刻后果之间的直接因果关联而进入更为复杂的系统中时,因果【Causal】关联会变得越来越弥散而难以把握。因此,影响评估也

是花费不菲的。

随着社会科学中更加重视系统理解的重要性，影响评估也变得越来越有价值，虽然实际执行起来仍然是有挑战性的。

——Michael Quinn Patton

（高勇译校）

参考文献

Lichfield, N. (1996). *Community impact evaluation*. London: University College Press.

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: Sage.

内隐量度（Implicit Measures）

内隐（implicit）过程和外显（explicit）过程的区分源于认知心理学的著作，但是目前在其他领域中也被普遍使用，如在社会心理学中。内隐过程是没有自觉意识、无意之中激发的，而外显过程是清醒的、有意的、可控的。对内隐过程的量度（即内隐量度）包括响应延迟程序（如内隐联想测验，Implicit Association Test）（Greenwald et al., 2002）；记忆任务（如过程分离程序，process-dissociation procedure）（Jacoby, Debner, & Hay, 2001）；生理反应（如皮电反应，galvanic skin response）；以及间接性问题（如涉及归因偏差的问题）。内隐量度包括的技术如此多样，如描述群体成员时使用更抽象还是更具体的语言的倾向，对于出现在自己姓名中的字母的偏好，在一对刺激中进行决策的反应时间。内隐联想测验（Implicit Association Test, IST）是一种流行的用于评估种族偏差的反应时间技术（Greenwald et al., 2002），在这种技术中，参与者首先要分别执行两项分类任务（如将名字分为黑人和白人两类，把单词分为正面和负面两类），然后这两项任务被合在一起执行。基于人们对彼此关联的一对刺激反应速度更快这一已经被证实的假定，对特定单词组合（如黑人姓名与正面特性的组合对黑人姓名与负面特性的组合）的决策速度就代表了内隐量度。与此不同，外显量度的典型例子就是传统的、直接的自陈式量度【Self-Report Measure】。

也许因为各种各样的内隐量度被设计出来处理反应的不同方面（如意识性或可控性），关注的也是不同的潜在过程或系统（如概念过程或感知过程），对于同一社会概念或态度的不同内隐量度之间只具有中等强度的关系，并且变化很大。此外，内隐量度的再测信度【Test-Retest Reliability】比外显量度稍低（Fazio & Olson, 2003）。

内隐量度和外显量度之间也只具有中等强度的关系，并且变化很大（Dovidio, Kawakami, & Beach, 2001）。内隐量度和外显量度之间不一致的潜在原因之一在于，它们在受到策略性控制影响方面有所差异。外显量度比内隐量度更为透明，具有更高的反应性【Reactivity】。因此，在具有社会敏感性的领域，内隐量度和外显量度之间的一致性更弱。内隐量度和外显量度之间不一致关系的另一种解释是，人们常常对于人和事物持有双重态度——一种是老旧的习惯性态度，另一种是新的自觉态度。因为内隐量度处理的是习惯性态度，外显量度评估的是新近生成的态度，如果双重态度的这两个成分有分歧，内隐量度和外显量度就可能并不相关。

总而言之，内隐量度代表了一系列形形色色的技术，这些技术在目标、实施、假定、解释上与直接而外显的量度有着根本不同。

此外，由于它们是新近出现的技术，其心理测量特性（信度【Reliability】与效度【Validity】）比许多已经成熟的外显量度（如偏见量表）来说更加缺少证明。但是，内隐量度的应用对于理解人类的记忆、认知和社会行为已经作出了相当大的贡献。

——John F. Dovidio

（高勇译校）

参考文献

Dovidio, J. F., Kawakami, K., & Beach, K. R. (2001). Implicit and explicit attitudes: Examination of the relationship between measures of intergroup bias. In R. Brown & S. L. Gaertner (Eds.), *Blackwell handbook of social psychology: Intergroup processes* (pp. 175-197). Malden, MA: Blackwell.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, *54*, 297-327.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, *109*, 3-25.

Jacoby, L. L., Debner, J. A., & Hay, J. F. (2001). Proactive interference, accessibility bias, and process dissociations: Valid subject reports of memory. *Journal of Experimental Psychology: Memory, Learning, and Cognition*, *27*, 686-700.

插补法（Imputation）

缺失值【Missing Data】在实践中很常见，通常使得数据分析变得棘手。因此，处理数据中的缺失值的一种通用方法就是插补法，即为每个缺失数据填充一个或多个合理的数值，以便产生一个或多个完整的数据集。原则上，每个这样的数据集都可以用标准方法进行分析，这些标准方法原本是适用于完整数据的。通常，先对缺失值进行插补然后再用标准的完全数据方法，要比设计用于直接分析不完整数据的专用统计技术更容易。例如，在社会调查中，在不同变量上出现缺失值很常见。收入信息最可能得不到，其次还有关于性习惯、卫生、宗教的问题。对观察到的不完整数据进行正确分析会非常复杂。因此，通常更容易的做法是插补法。

插补法的另一个重要优势是在需要产生一个面向公众使用的数据的场景下，因为插补者会使用到那些不适于向公众公布的备用的保密信息和细节信息。除此之外，数据插补法使得所有使用者都以相同方式解决缺失值问题，这确保了所有使用者的分析是一致的。

插补法也能够容纳那些在标准统计软件包进行的分析中通常被删除掉的已观察数据。这些软件包通常用整条删除【Listwise Deletion】法处理不完整数据，也就是说，删除所有至少有一个缺失项的案例。这种做法在统计上是低效的，通常会导致有很大偏误的推断。特别是在多元分析中，“整条删除”会大量减少可用数据，使得剩余数据不再对目标总体具有代表性。只基于完整观察案例的多元分析通常会损失相当大比例的数据（30%甚至更多），这至少减少了估计的精确性。

除此之外，只基于完整案例进行推断有一个隐含的假设，即缺失数据是完全随机缺失的（Missing Completely at Random，MCAR），通常这一假设并不成立。在社会调查中，特殊的社会经济群体或少数民族通常会产

生更多的缺失值。在这种情况下,如果缺失与否可以用已观察变量而非缺失变量来解释,缺失数据就被称为是随机缺失的(missing at random,MAR),就可以基于已观察变量来进行有效插补。作出如下假定通常是合理的,即影响回答行为差异的变量,如性别、年龄组、生活状况、社会地位等,都已经被观察到。

如果缺失与否取决于缺失变量自身,那么数据就是非随机缺失的(missing not at random,MNAR)。(正式定义请参见Little & Rubin,2002)收入情况的回答就可能是这种情况,高收入者更可能不回答,即使在已观察变量的类别给定下(如性别、居住地等)也是如此。但是,即使在这类情况下,单一插补法和多重插补法仍然有助于减少由于不回答产生的偏差。

单一插补法

插补法通常用于题项无回答【Nonresponse】,但是它也可以用于单位无回答。许多直觉上很吸引人的方法已经发展出来用于单一插补,即对每个缺失数据插补一个数值。一种略为粗糙的方法用变量的总样本均值或者案例分组后的分组样本均值(如按性别分组)来代替此变量的每个缺失值。在回归插补法中,用完整案例来估计有缺失值的变量对其他已观察变量的回归【Regression】方程。然后使用得到的预测方程来估计条件均值,用以插补每个缺失值。在随机回归插补法中,一个随机残差加入回归预测中,而随机误差的方差与回归中估计出的残方差相同。

另一种常见的程序是热卡插补法(hot deck imputation),即用所谓的供体案例(donor case)的已观察值来代替不完整案例的缺失值。简单的热卡插补法先要基于某些完整案例和不完整案例均已观察到的变量(如性别和居住地)的交互分类来定义插补单元格(imputation cells)。然后数值已经被观察到的那些案例与有缺失值的每个案例要进行配对,构成供体源,随机选择的一个供体案例的已观察值就被用来移植(插补)代替缺失值。

在最近相邻插补法中,已观察变量的距离(度量)被用来定义最佳的供体案例。当距离是基于案例在欲插补变量的预测值上的差异时,这种插补程序称为预测均值匹配法(predictive mean matching)。在过去十年间,在发展有用的热卡插补程序和最近相邻程序上已经取得了巨大的进展。

利特尔(Little,1988)详尽地讨论了产生插补值中的问题。对于插补法来说,一个主要的考虑是,在考虑了所有已观察取值之后,应该用随机抽取而不是用缺失值的最佳预测值。用均值、分组均值、回归预测值等点估计来取代缺失值,易于扭曲数据中非线性的那些估计量,如方差【Variance】、协方差【Covariance】、相关系数【Correlation】。总之,均值插补法有损于有效估计,更有损于有效推论(如置信区间【Confidence Interval】)。随机回归插补法和热卡程序对点估计而言还不错,但对于方差估计就需要进行特别修正。针对热卡法和最近相邻技术的方差估计方法已经发展出来,但是只限于特别的双变量统计量;对于此问题最新的讨论,请参见 Groves, Dillman, Eltinge, & Little(2002)。

多重插补法

为每个缺失值插补一个数值,然后用适用于完整数据的标准技术来分析此完整数据,上述做法通常会导致标准误差估计过小,置信区间过窄,而 p 值过分显著;即使绝对正确地执行了插补法的建模,上述问题依然存在。鲁宾在 1978 年引入了多重插补法

(multiple imputation, MI)并进行了详尽讨论(Rubin, 1987),这种方法可以在保有插补法优势的同时,使得数据分析者对于不确定性进行有效评估。MI 是一种用 $m>1$ 个模拟值代替缺失值的蒙特卡洛技术,模拟值根据给定已观察数据条件下真实值出现可能性的概率分布给出。m 通常很小,如 $m=5$。每个插补后的数据(即完整数据)首先用标准方法进行分析,然后再将结论汇总起来产生反映缺失数据不确定性的估计量和置信区间。

多重插补法的理论动机是贝叶斯学派的,但得出的多重插补推断从频率学派的观点来看也是成立的。正式地,贝叶斯 MI 要求进行在给定已观察数据的前提下,从缺失值的后验预测分布中进行独立的随机插取:

$$f(y_{mis}|y_{obs}) = \int f(y_{mis},\theta|y_{obs})\,d\theta = \int f(y_{mis}|y_{obs},\theta)f(\theta|y_{obs})\,d\theta$$

(为简便起见,这里用通俗的"分布"来代替严格的"概率密度函数"。)直接从 $f(y_{mis}|y_{obs})$ 中抽取通常很困难,所以对于 m 次抽取中的每一次都要有以下的两步骤程序:

(1)根据已观察数据的后验分布 $f(\theta|y_{obs})$ 来抽取参数 θ。

(2)在已观察数据和从步骤 1 中抽取出的参数值给定的条件下,根据其条件预测分布 $f(y_{mis}|y_{obs},\theta)$ 来抽取缺失值 Y_{mis}。

对于很多模型,由于使用的完整数据模型具有 iid 性质,步骤 2 中的条件预测分布 $f(y_{mis}|y_{obs},\theta)$ 是相对直截了当的。然而,相应的已观察数据的后验分布 $f(\theta|y_{obs}) = f(y_{obs}|\theta)f(\theta)/f(y_{obs})$ 通常很难得到,特别是当数据是多元的,而且缺失模式比较复杂时。因此,已观察数据的后验分布通常不是易于从中产生随机抽取的标准分布。但是,直接的迭代插补程序已经发展出来,可以通过马尔科夫链蒙特卡洛方法【Markov Chain Monte Carlo Methods】来进行多重插补。谢弗(Schafer,1997)深入讨论了许多例子。这些包括了诸如 Gibbs 抽样【Gibbs Sampling】的数据扩充程序(Tanner & Wong, 1987)可以产生一个随机系列$\{(y_{mis}^{(t)},\theta^{(t)}),t=1,2,\cdots\}$,其平稳分布为$f(y_{mis},\theta|y_{obs})$。具而言之,产生了如下迭代抽样方案。给定当前对于参数的猜测值 $\theta^{(t)}$,首先从条件预测分布 $f(y_{mis}|y_{obs},\theta^{(t)})$ 中抽取一个 $y_{mis}^{(t+1)}$ 作为缺失值。然后,基于 $y_{mis}^{(t+1)}$ 从其完整数据后验分布 $f(\theta|y_{obs},y_{mis}^{(t+1)})$ 中抽取新的 θ 值 $\theta^{(t+1)}$。假定 t 足够大时,从这些马尔科夫链中进行的 m 次独立抽取就可以用来作为从其后验测验分布 $f(y_{mis}|y_{obs})$ 得到的 Y_{mis} 的多重插补值。

现在考虑估计一个未知量 Q 的任务。在本节余下的部分中,未知量 Q 是从多重插补后的数据集中估计出的,它不同于插补模型中使用的参数 θ。虽然 Q 可以是 θ 的显函数,但多重插补法的优势之一就是并不一定需要如此。事实上,Q 甚至可以是多重插补后数据集的分析者选定的模型中的参数,而分析者的模型可以与插补者的模型不同。只要两个模型不是完全不相容,或者缺失信息比例不是太高,基于多重插补数据之上的推论仍然是大体成立的。鲁宾(Rubin, 1987)和谢弗(Schafer, 1997, Chap.4)及其参考文献中更深入地讨论了 Q 和 θ 之间的区分。

MI 原则假定完整数据统计量 $\hat{Q}$ 及其方差估计 $\hat{V}(\hat{Q})$ 可以视为 Q 的近似完整数据后验均值和方差,即基于合适的完整数据模型和先验分布而得到的 $\hat{Q}\approx E(Q|y_{obs},y_{mis})$ 和 $\hat{V}(\hat{Q})\approx V(Q|y_{obs},y_{mis})$。除此之外,使用完整数据基于大样本正态近似的检验和区间估计同样成立:

$$(\hat{Q}-Q)/\sqrt{\hat{V}(\hat{Q})} \sim N(0,1)$$

注意,从反转费希尔信息矩阵中得到的通常的最大似然估计【Maximum Likelihood Estimation】及其渐进方差一般满足上述假定。有时,需要将估计量 $\hat{Q}$ 转换为可用于正态近似的尺度;有关其他实践中的问题(请参见 Rubin,1987;Schafer,1997;Little & Rubin,2002)。

在创建了 m 个经过插补的数据集之后,m 个完整数据的统计量 $\hat{Q}^{(i)}(i=1,\cdots,m)$ 及其方差估计 $\hat{V}(\hat{\theta}^{(i)})$ 就易于计算了。最终的对于量 Q 的 MI 点估计 $\hat{Q}_{MI}$ 就是其平均值:

$$\hat{Q}_{MI}=\frac{1}{m}\sum_{i=1}^{m}\hat{Q}^{(i)}$$

根据方差分析【Analysis of Variance】原则,总方差估计 T 以如下方式计算:

"插补间"方差

$$B=\frac{1}{m-1}\sum_{i=1}^{m}(\hat{Q}^{(i)}-\hat{Q}_{MI})^2$$

"插补内"方差

$$W=\frac{1}{m}\sum_{i=1}^{m}\hat{V}(\hat{Q}^{(i)})$$

"总方差"

$$T=W+\left(1+\frac{1}{m}\right)B$$

修正量$\left(1+\frac{1}{m}\right)$反映了 Q 的估计是从有限次(m)插补中得到的。检验及双侧区间估计可以基于以下近似获得:$(\hat{Q}_{MI}-Q)/\sqrt{T}\sim t_v$,其中 t_v 表示某一自由度下的学生 t 分布。

$$v=(m-1)\left(1+\frac{W}{(1+m^{-1})B}\right)^2$$

巴纳德和鲁宾(Barnard and Rubin,1999)放宽了上述完整数据区间估计及检验时采用 t 分布的正态参照分布假定,他们得出了相应的 MI 推断的自由度以替代前面给出的 v 的公式。除此之外,还可以采用其他方法来综合向量估计和协方差矩阵、p 值、似然比统计量【Likelihood Ratio】(参见Little & Rubin, 2002)。

比起基于单一插补法进行的估计来说,多重插补法的区间估计会得到一个较大但有效的区间,因为这个区间考虑到了缺失值的不确定性和模拟误差而加宽了。即使插补次数不多(比如说,$m=3$ 或 4),得到的推断一般而言也接近于其名义涵盖范围或显著水平。热卡程序或随机回归插补也可以用来产生多重插补,但是这些不是从贝叶斯框架中得出的程序通常在鲁宾(Rubin,1987)定义的意义上并不是正规而适当的,因为由于估计插补模型的参数而导致的额外不确定性并非反映出来。

准确的 MI 方法正确地反映了抽样的变异性;也就是说,如果插补模型本身正确,从频率学派的观点来看,得到的多重插补推论也是正确的。随机回归插补法可以改造得更准确,办法是遵循前述两阶段程序从后验预测分布中抽取 Y_{mis},首先抽取回归模型参数(第 1 步),然后在通常插补程序中使用抽取到的参数(第 2 步)。鲁宾和申克尔(Rubin and Schenker,1986)建议使用自助抽样法改造热卡插补法,得出的程序被称为近似贝叶斯自助法。这种两阶段程序首先从每个供体源(donor pool)中抽取一个自助样本,然后从自助样本中随机抽取插补值;这是一种抽取 Y_{mis} 的两阶段程序的非参数

类比。因此,近似准确的多重插补法常常可以通过直接使用标准插补程序得到。

计算问题

各种单一插补技术,如均值插补、条件均值插补、回归插补,在诸如 SPSS 这类商业统计软件包中都可以实现。随着计算能力增加,越来越多的多重插补法将被实现,使得多重插补程序对产生多重插补数据集和分析此类数据集都变得非常容易实现。

还有一些免费的程序和例程,如独立 Windows 程序 NORM,或者 S-PLUS 中的程序包 NORM,CAT,MIX,PAN,MICE(现在 R 用户也可以使用这些程序包)。NORM 对连续数据使用正态模型,CAT 对类别数据使用对数线性模型。MIX 基于一般位置模型(General Location Model)处理兼有类别和连续的数据。PAN 用于处理面板数据。S-PLUS 6 中的新缺失值程序包已经包括了上述模型并进行了简化,在多重插补之后整合了多个完整数据分析的结果。

除此之外,还有免费的 SAS 可调用的应用程序 IVEware,还有基于正态分布的 SAS 程序 PROC MI 和 PROC MIANALYZE。PROC MI 提供了参数和非参数的回归插补方法,以及多元正态模型。MICE 和 IVEware 都是非常灵活的工具,可以使用链式方程为各种类型的变量生成多元插补值。AMELIA(有免费的 Windows 版本和 Gauss 版本)是专门设计出来以满足某些社会科学家的需求的。

最后,SOLAS for Missing Data Analysis 3.0 是由统计解决方案公司(Statistical Solutions Limited)提供的一套商业 Windows 程序(链接和相关细节,请参看 Horton & Lipsitz,2001)。

总结与扩展

插补法使调查者可以立刻使用所有的可得技术来分析完整数据。我们对此进行了总论,并对产生插补值中的模型问题进行了简要讨论,还讨论了如何分析多个插补后的数据集以得到有效推断。

对于社会科学而言,MI 的另外一个优势可能在于将较长问卷分为合适的几部分,以便回答人只需要填答较短的问卷,这保证了更高的回答率、更好的数据质量,而且还降低了成本。由此生成的缺失数据可能设计成为完全随机缺失(MCAR)或随机缺失(MAR)的,然后可以进行多重插补,得到的结果也很好;有关其基本思路,请参见 Raghunathan & Grizzle(1995)。

虽然贝叶斯方法常因其需要选择先验分布而为人诟病,但在识别问题发生时这一点本身也会成为一种优势。例如,在不同的先验情景下的缺失值多重插补可以进行敏感性分析(sensitivity analysis)[有关其在统计匹配(statistical matching)中的应用,请参见 Rubin,1986;Rässler,2002]。

最后,经验证据表明,随机缺失(MAR)的多重插补对违反 MAR 假定通常是相当稳健的。即使 MAR 假定是错误的,使用多重插补策略对估计值和标准差也只有很小的影响。只有当非随机缺失(MNAR)是严重隐患,缺失信息的比例相当大时,才有必要对数据和缺失值进行联合建模。除此之外,由于缺失值不可观察,数据并不存在直接证据来说明 MAR 假定。因此,更有帮助的办法是考虑多个替代模型,然后考察得到的推断的敏感性。

因此,多重插补程序看起来是考虑缺失问题和利用所有有效的宝贵信息的最佳可行办法。总之,要使随后的分析在统计上有效,关键一点是要使用多重插补而非单一

插补。

——Susanne Rässler
Donald B. Rubin
Nathaniel Schenker
（高勇译校）

作者注释:本词条表达的观点并不一定代表美国政府的观点,申克尔作为合著者并不意味着官方对其陈述的认可,甚至其陈述可能与美国政府官方立场相左。

参考文献

Barnard, J., & Rubin, D. B. (1999). Small-sample degrees of freedom with multiple imputation. *Biometrika*, *86*, 948-955.

Groves, R. M., Dillman, D. A., Eltinge, J. L., & Little, R. J. A. (2002). *Survey nonresponse*. New York: Wiley.

Horton, N. J., & Lipsitz, S. R. (2001). Multiple imputation in practice: Comparison of software packages for regression models with missing variables. *The American Statistician*, *55*, 244-254.

Little, R. J. A. (1988). Missing data adjustments in large surveys. *Journal of Business and Economic Statistics*, *6*, 287-301.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. New York: Wiley.

Raghunathan, T. E., & Grizzle, J. E. (1995). A split questionnaire survey design. *Journal of the American Statistical Association*, *90*, 54-63.

Rässler, S. (2002). Statistical matching: A frequentist theory, practical applications, and alternative Bayesian approaches. *Lecture Notes in Statistics*, *168*. New York: Springer.

Rubin, D. B. (1986). Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business and Economic Statistics*, *4*, 87-95.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Rubin, D. B., & Schenker, N. (1986). Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association*, *81*, 366-374.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman and Hall.

Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association*, *82*, 528-550.

独立(Independence)

社会科学家们通常对于此类经验问题很感兴趣:“在大都市地区居住和面临更高犯罪率是否有关”,或者“一年挣到 50 000 元以上与获得大学文凭是否有关”。我们用统计独立(independence)或者相依(dependence)这样的概念来回答这些问题。如果某一事件的发生并不影响另一事件的发生概率【Probability】,两个事件就被称为统计上彼此独立。这一普遍性的思想可以扩展到任何数量的彼此无关的独立事件。相反,如果某一事件的发生影响另一事件的发生可能性,这两个事件就被称为是相关的,或者统计上是相依的。

如果两个事件彼此独立,某一事件发生的事实就不会影响到另一事件发生的概率。假定事件 A 的概率[即 $P(A)$]和事件 B 的概率[即 $P(B)$]都不为零。两个事件之间的统计独立意味着,B 给定时 A 的条件概率[即 $P(A|B)$]与 A 的边缘概率(或非条件概率)[即 $P(A)$]完全相同。也就是说,如果 A 和 B 是独立的,那么 $P(A|B)=P(A)$。类似地,如果两个事件彼此独立,可以期望 A 给定时 B 的条件概率[即 $P(B|A)$]与 B 的发生概率[即 $P(B)$]完全相同。换言之,如果

A 和 B 彼此独立，那么 $P(B|A)=P(B)$ 且 $P(A|B)=P(A)$。

统计独立的通常定义是，当且仅当事件 A 和事件 B 的联合概率[即 $P(A\cap B)$]等于它们各自边缘概率(或非条件概率)之积时，这两个事件彼此独立。试考虑两个事件 A 和 B。我们已知 $P(A|B)=P(A)$，那么因为 $P(A|B)=\frac{P(A\cap B)}{P(B)}$，我们可以在 $P(B)\neq 0$ 的假定下得到 $P(A\cap B)=P(A)\times P(B)$。类似地，$P(B|A)=P(B)$ 也可以用相同方式得出同样的结论 $P(B\cap A)=P(A)\times P(B)$。彼此独立的定义可以扩展到多个事件中。具体来说，事件 $A,B,C,\cdots,Z$ 的概率均不为零，它们当且仅当如下条件成立时彼此独立：

$$P(A\cap B\cap C\cap\cdots\cap Z)=P(A)\times P(B)\times P(C)\times\cdots\times P(Z)$$

假定我们抛一个公平的六面骰子两次。事件 A 为出现奇数(出现 1,3,5)，事件 B 为出现偶数(出现 2,4,6)。我们知道，$P(A)=3/6=1/2$，$P(B)=3/6=1/2$。那么抛第一次出现奇数(事件 A)且第二次出现偶数(事件 B)的概率为 $P(A\cap B)=P(A)\times P(B)=1/2\times 1/2=1/4$，因为这两次抛骰子都是彼此独立的。类似地，抛一个没有偏差的硬币三次，得到三次反面的概率就是 $P(T_1\cap T_2\cap T_3)=P(T_1)\times P(T_2)\times P(T_3)=1/2\times 1/2\times 1/2=1/8$。

表 1　人口交互分类显示的统计相关性

	党　派			
种　族	民主党	独立党	共和党	总和
白人	400	100	600	1 100
非白人	100	20	260	380
总和	500	120	860	1 480

考虑一个统计独立在其中不成立的例子。表 1 是一个列联表【Contingency Table】，显示的是一个纯粹假想出来的社区中的人口交互分类情况。如果从这个总体中随机抽取一个居民，这个居民是白人且是共和党的概率是多少？令事件 A 为我们抽到一个共和党人，B 为我们抽到一个白人。因为共有 860 名共和党人，所以 $P(A)=860/1\,480$，概率大约是 0.581。类似地，抽到一个白人的概率为 $P(B)=1\,100/1\,480$，大约是 0.743。在 1 100 名白人中，有 600 人认同自己是共和党人。因此，如果我们已知这个居民是白人，那么他是共和党人的概率就是 600/1 100，大约是 0.545。类似地，如果我们已知这个居民是共和党人，那么他是白人的概率就是 600/860，大约是 0.698。因此我们有 $P(A)=0.581$，$P(B)=0.743$，$P(A|B)=0.545$，$P(B|A)=0.698$。可以看出，在这个例子中，$P(A\cap B)\neq P(A)\times P(B)$，因此我们得到：

$P(A\cap B)=P(A)\times P(B|A)=0.581\times 0.698\cong 0.406$

或 $P(A\cap B)=P(B)\times P(A|B)=0.743\times 0.545\cong 0.405$

但是，$P(A)\times P(B)=0.581\times 0.743=0.432$，我们发现 A 和 B 的联合概率 $P(A\cap B)$ 并不等于它们各自概率的乘积。因此这两个事件是相关的。用来检验变量间独立性的统计方法取决于数据的性质[更多关于独立性的统计检验的信息请参看 Agresti & Finlay(1997)，Agresti(2002)，Johnson & Bhattacharyya(2001)]。

——Cheng-Lung Wang

(高勇译校)

参考文献

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley Interscience.

Agresti, A., & Finlay, B. (1997). *Statistical methods*

for the social sciences. Upper Saddle River, NJ: Prentice Hall.

Johnson, R. A., & Bhattacharyya, G. K. (2001). *Statistics: Principles and methods* (4th ed.). New York: Wiley.

自变量(Independent Variable)

自变量就是被认为对另一个被称为因变量【Dependent Variable】的变量施加因果影响的任何变量。什么可以被视为自变量,实验研究和非实验研究对此的理念有所不同,请参看自变量(实验研究中)【Independent Variable (In Eexperimental Research)】和自变量(非实验研究中)【Independent Variable (In Nonexperimental Research)】。

——Alan Bryman

(高勇译校)

自变量(实验研究中)(Independent Variable [in Experimental Research])

每一个真正的实验的标志性特点就是有自变量(Independent Variable, IV),即一个由研究者控制的可调节或者可变化的实验特征。此类变量被称为自变量,因为它们独立于它们在自然条件中的时空变化来源。研究者可以决定是否施加这个自变量于某一特定参与者身上,或者某些参与人群中,可以决定这个变量的程度或强度。对自变量的有控制的变化,被称为实验处理(treatments),或者实验操控(manipulations)。自变量是否存在或者自变量的不同数量被称为处理的不同水平。处理水平可以是定量的(如某种药物的5,10,35 mg),也可是定类的(参与者被分配到工作小组A,B或C)。自变量可以与因子设计【Factor Design】结合起来,然后估计它们的交互影响。在实验室实验【Laboratory Experiment】中,自变量非常关键,因为它构成了因果推断的基础。被随机分配到不同实验条件下的参与者被认为是在起始阶段完全相同的。接受了自变量不同水平的参与者在因变量【Dependent Variable】上的处理后,差异被归因于自变量。

科拉诺和布鲁尔(Crano & Brewer, 2002)区分了三种类型的自变量:社会的、环境的、指令的处理。社会处理取决于实验中人的行动,这些人通常是实验者雇用的演员,他们的行为是预先规定好的或者受控制的。在阿施(Asch, 1951)的经典遵从研究中,天真的参与者做出了一系列与其他人一致的比较性评价。事实上,其他人都不是天真的,而是一直在把参与者引入歧途的共谋者。在一种情况下,会有一个共谋者给出与其同伙不一致的结论,这就破坏了共谋者之间的(错误的)一致意见。在意见一致和不一致两组之间的差异,就被解释为是受控制的共谋者的反应一致程度差异引起的。

环境处理指的是对于物理场景的某些特质进行系统操控。在一项态度变化实验中,所有的参与者都接受相同的劝服性交流,但是有人是在存在干扰的环境下接受这些信息的,有着实验故意制造出来的噪声和混乱。在正常条件下的参与者和存在干扰的环境下的参与者,这两组受到信息影响的程度就可以进行因果解释,只要参与者是被随机地分配到两组之中的,并且实验者控制

着干扰因素的存在。

指令操控取决于提供给参与者的指令差异。例如,赞纳和库珀(Zanna & Cooper,1974)给参与者一片药,然后告诉其中一些人他们会感到紧张不安,告诉另一些人药片会有助于放松。事实上,药片含有一种惰性物质,根本没有任何药理效果。参与者在随后产生的判断差异就被归因于指令操控(即自变量)的差异。

对于严格的实验主义者来说,区分参与者的因素(如性别、宗教信仰、智商、人格因素)和其他不在研究者操纵范围之内的变量(如洛杉矶的自杀率)不能被认为是自变量的,也不能进行因果解释。但是,在一些研究传统中,不在实验操控下的变量有时也被认为是原因。例如,一个社会中在时间点 1 上的结婚率和时间点 2 上的自杀率存在高度相关,人们可能会想到,婚姻增加了人的满足感,进而影响了自杀发生的可能性。这可以解释为,婚姻引起了满足,这影响了自杀率的发生可能性。相反的结论是站不住脚的:时间点 2 上的自杀率不可能影响先前的结婚率。但是,这种结论通常也应当是试探性的,因为可能存在第三个没有测量的变量才是这一表面上的关系的真正原因。在现在这个例子中,经济条件可能是真正的原因,它在时间点 1 影响了结婚率,在时间点 2 影响了自杀率。由于存在这种第三个变量才是真正原因的可能性,基于相关结果的因果推论至多被认为是试探性的。在理想的实验情境中,这些外生影响被控制住了,因此不能用它们的影响来解释接受了不同水平的变量处理的各组间的差异。

——William D. Crano

(高勇译校)

参考文献

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership, and men* (pp. 177-190). Pittsburgh, PA: Carnegie Press.

Crano, W. D., & Brewer, M. B. (2002). *Principles and methods of social research* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Zanna, M. P., & Cooper, J. (1974). Dissonance and the pill: An attribution approach to studying the arousal properties of dissonance. *Journal of Personality and Social Psychology*, *29*, 703-709.

自变量(非实验研究中)(Independent Variable [in Nonexperimental Research])

自变量也被称为解释变量【Explanatory Variable】或者输入变量。自变量,或者更准确地说,自变量对于因变量【Dependent Variable】的影响,是研究者在典型的定量社会科学研究中分析的对象。严格而言,自变量是研究者能够操控的变量,如其在实验【Experiment】中那样。但是,在非实验研究中,变量是无法操控的,因为有时候研究者并不清楚研究的多个变量之中哪一个可以被视为自变量【Independent Variable】。有时候,研究者宽泛地使用这一术语,来指涉回归【Regression】方程中处于右端的所有变量。但是有必要讨论在不同的研究设计【Research Design】中自变量的内在含义。如果数据是从横截面设计【Cross-Sectional Design】中得到的,所有变量信息都是一起获取的,没有变量是被操控的。这意味着,哪个变量对其他变量有因果影响并非显而易见。在社会学、政治学、经济学、地理学等学科中的变量不被操控,主要有以下几个原

因:某些变量不可能被操控(如民族和性别);对某些变量操控并不现实(如人们居住的地区);或者操控在伦理上或政治上是不可接受的(如引发贫困的因素)。有一些此类变量(如种族、民族、性别、出生国等),在分析中被视为"固定"的。与实验研究中可操控性的定义相反,固定变量是根本不可操控的,但是在非实验研究中却是可以被安全地视为和处理为"自变量"。

在纵贯性设计中,如定群设计【Panel Design】,问题会变得更为复杂,因为虽然变量仍然不能操控,但是数据是在不同时点收集的,这使研究者可以对变量在时间上的先后和因果关系的方向进行经验性的分析。严格的实验主义者仍然会坚称,操控的缺乏会使得因果关系的方向令人怀疑。为了解决这一困难,定群数据的体制者求助于所谓的交叉滞后定群分析(cross-lagged panel analysis),来评估在两个或多个时点上测量的变量对中的因果关系方向。此类分析中自变量和因变量的区分在某种程度上变模糊了,因为在第一个回归方程中被认为因果上在先的变量被视为自变量,但是在第二个回归方程中就被视为因变量。反过来,在第一个回归方程中被认为因果上在后的变量被视为因变量,但是在第二个方程中被视为自变量。从此类分析来看,自变量(independent variable)这一术语就显得有些武断;要紧的是在非实验研究中进行清晰的思考和仔细的分析。

对于从非实验研究中收集的变量运作一个回归方程,然后就进行因果关系推断,这已经众所周知是错误的,但是我们在这些研究中如何才能获得自变量呢?使用非实验研究设计(如调查【Survey】设计)的研究者必须进行因果推断(causal inference)来探寻出自变量。本质上,这个过程既包括从我们对社会世界本质的理解中进行常识性推断,也包括从关注领域中的现有理论、研究和分析技术中进行推断。这一因果推断过程存在于诸如因果建模【Causal Modeling】和定群分析【Panel Analysis】这类方法当中,交叉滞后定群分析也与此有关。

举一个简单的例子:众所周知,年龄与投票行为是关联的。这是否意味着,我们所能说的只是这两个变量是相关的?因为我们的投票方式不可能对我们的年龄有任何影响,因此认为年龄是自变量看起来更为合乎情理。当然,可能存在中介变量【Intervening Variable】或者调节变量【Moderating Variable】在起作用,它影响着年龄和投票行为之间的因果路径,但是此类变量可以进行因果推断这一基本点仍然成立。

如果因果方向不是那么显而易见,那么进行因果推断就变得更为困难了。例如,如果我们从一个大公司的调查中发现雇员的组织忠诚水平和工作满意度是相关的,就很难确定因果关系的方向,虽然与这些变量相关的既有发现和理论可能会提供给我们一些有用的线索。在此类情形中,面板研究通常会是分解原因和结果的较好工具。

——Alan Bryman
Tim Futing Liao
(高勇译校)

参考文献

Davis, J. A. (1985). *The logic of causal order* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-055). Beverly Hills, CA: Sage.

Rosenberg, M. (1968). *The logic of survey analysis*. New York: Basic Books.

深度访谈(In-Depth Interview)

深度访谈是一个术语,指涉包括无结构化访谈【Unstructured Interview】,半结构化访谈【Semistructured Interview】,有时也包括其他类型的定性研究中的访谈【Interview in Qualitative Research】,如叙事访谈【Narrative Interview】和生活故事访谈【Life Story Interview】。

——Alan Bryman

(高勇译校)

指数(Index)

指数是对一个抽象的理论构念【Construct】的测度,它将这个构念的两个或多个指标(indicators)综合起来构成单一的汇总得分。就这一点而言,量表【Scale】就是一种指数,事实上这两个词有时可以互换使用。但是,量表以数据中属性模式为基础对个体进行排列,而指数只是几项被称为子项(items)的指标的加总合成。因此,量表可以利用数据中属性间存在的任何强度结构,而指数只是假定所有子项都同样程度地反映了潜在的构念,因此只要加总此人在每个子项上的得分就可以反映这个构念。

指数被政府机构和社会科学界广泛使用,比量表使用更为频繁。例如消费者价格指数测量了经济中不同领域(如住房、洗车、地产、消费品等)的价格。所有这些东西都会影响到消费者为商品和服务而支付的平均价格,因此都被包含到此指数当中。这个例子表明了为什么指数一般而言优于单个指标。汽油价格在某个时段里可能暴涨,但是其他东西的价格仍然保持稳定。在其他东西价格不变时,只是基于上涨的汽油价格来建立消费者价格指数是不准确且具有误导性的。

此外,指数不一定是单一维度的。例如,政治学家使用一个七级指数测量党派认同。这个指数沿一个连续谱如下排列:坚定的民主党,一般的民主党,偏向民主党的独立人士;独立人士;偏向共和党的独立人士;一般的共和党,坚定的共和党。很容易看出,这个指数包括了两个不同的维度。第一个是方向维度,从民主党到独立人士到共和党。但是,还有一个强度维度,从坚定到一般、独立、一般,再回到坚定。

政府常常使用由官方统计数据构建的指数。例如,消费者价格指数被用来测量消费者支付的价格水平,FBI 的犯罪指数是七项所谓的指数犯罪率(index crime rates)的加总,它是美国犯罪情况的整体指标。

——Edward G. Carmines

James Woods

(高勇译校)

参考文献

Carmines, E. G., & McIver, J. P. (1981). *Unidimensional scaling* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-024). Beverly Hills, CA: Sage.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

指标(Indicator)

参见指数【Index】。

归纳(Induction)

在逻辑上,归纳就是从特殊陈述到普遍陈述的过程。在社会科学中,这种逻辑被用来从资料中生成理论【Theory】。许多具体实例被用来产生一个普遍结论,这个普遍结论的论断范围超出了其立基的证据。理论包括关于现象间联系的普遍性概括【Generalization】。例如,许多来自破裂家庭的少年罪犯的观察实例被用来得出结论:"所有的少年罪犯都是来自破裂家庭。"

这是对科学家的工作方式的流行看法,他们进行细致而客观的观察和测量,然后仔细而精确地分析资料。这也是与实证主义【Positivism】相关联的科学逻辑,其反面是与证伪主义【Falsificationism】相关联的演绎【Deduction】。

归纳推理的早期形式要归功于亚里士多德及其弟子,这也被称为枚举归纳(enumerative induction)或单纯归纳(naïve induction)。之后在17世纪,弗兰西斯·培根倡导将归纳视为唯一的科学方法。他想要的方法是要能够"从头脑中清除所有的预设与偏见,带着全新的眼光去阅读自然的书"(O'Hear, 1989, p.16)。培根批评亚里士多德的方法,认为不能仅仅通过从观察中普遍化来累积知识,还有必要关注那些负面的实例。他的方法是一种"清除性的(eliminative)"归纳法,要求对存在和不存在要考察的现象的实例进行比较。两个世纪之后,约翰·斯图尔特·密尔在他的求同法和差异法中将培根的思想精细化了。在求同法中,两个现象存在的实例中需要除了一个特点之外完全不同;在差异法中,两种情形被要求除了一个特征之外完全相同,而现象存在与否就取决于这个特征。

归纳逻辑的倡导者要求必须使用客观的【Objectivity】方法,要认真而系统地收集资料,不要有任何预先的理念来指引他们的选择。概括是从资料中逻辑性地产生的。一旦形成了概括,可以收集更多资料来佐证或验证这一说法。概括出来的模式也可以用来解释新的观察实例。例如,如果少年罪犯和破裂家庭之间的关系的概括可以被视为普遍性的,那么少年罪犯的更多案例就可以从家庭背景中得到解释,或者可以从家庭背景中去进行犯罪预测。这被称为模式解释(pattern explanation)。

许多科学哲学家都严格地批评和拒斥了归纳。如下论述都引发了许多辩驳:先入之见可以被置之一旁来得到客观的观察;没有理念来指导选择也能够进行"切题的"观察;归纳逻辑有能力从资料中自动地产生概括;基于有限观察可以得到普遍性的概括;只要发现规律性就足以产生解释(批判性的评论,请参见 Blaikie, 1993; Chalmers, 1982; O'Hear, 1989)。演绎提供了另一种替代方法。

在社会科学中,涂尔干(Durkheim, 1964)最早在其《社会学方法准则》(*Rules of Sociological Method*)中倡导使用归纳法,并在《自杀论》(*Suicide*)研究中将归纳法应用于实践。尽管存在缺陷,在关注回答"是什

么"研究问题【Research Question】的描述性社会研究中,经过修正之后的归纳法仍然非常重要(参看 Blaikie, 2000)。

——Norman Blaikie

(高勇译校)

参考文献

Blaikie, N. (1993). *Approaches to social enquiry*. Cambridge, UK: Polity.

Blaikie, N. (2000). *Designing social research: The logic of anticipation*. Cambridge, UK: Polity Press.

Chalmers, A. F. (1982). *What is this thing called science?* St. Lucia: University of Queensland Press.

Durkheim, E. (1964). *Rules of sociological method*. Glencoe, IL: Free Press.

O'Hear, A. (1989). *An introduction to the philosophy of science*. Oxford, UK: Clarendon.

不平等测量(Inequality Measurement)

个体在许多方面存在差异。因此,群体的明显特征就是在个体存在差异的那些方面具有不同程度的不平等。个体间差异的几个重要方面包括收入、财富、权力、预期寿命、数学能力。个体间的不平等存在于上述每个维度中,不同的总体在个体间的各种特征上有着不同分布。通常一个给定总体在子群体间也存在不平等:男性与女性;乡村与城市;白人与黑人。确定和测量社会中不同维度的不平等,通常都很重要。出于公共政策原因,我们对于代表或影响个体福利、权利、生活前景差异的社会不平等尤其感兴趣。

我们如何才能测量一个总体【Population】中的不平等程度呢?取决于这一因素的分布在总体中是否近似地服从正态分布【Normal Distribution】,存在几种重要的不同方法。如果这一因素近似地服从正态分布,那么不平等测量就可以通过对这一因素在总体中的均值【Mean】和标准差【Standard Deviation】分析来完成。在这种方法中,我们关注总体在这一特征(收入、预期寿命、体重、成绩)上的描述性统计量:得分的分布形状、总体的均值和中位数【Median】得分,围绕均值得分的标准差。变量【Variable】在总体中的标准差提供了这一特征在总体中的散布程度的客观测量。我们可以用此类统计测量对于单个总体中的不同群体间的不平等程度得出结论("女性工程师的平均工资只有男性工程师的70%");也可以对于不同总体中的不平等程度得出结论("美国成人体重的方差是均值的 20%,而在法国,方差只是均值的12%"),以此类推。

另一种不同的测量方法用于一类重要的不平等——与总体中资源分布有关的不平等。财富和收入的分布,以及其他重要社会资源的分布,通常在统计上都不服从正态分布;通常这类分布都是严重偏向【Skewed】于低端的(也就是说大量的个体只拥有少量的分配)。在此,我们对总体中从最低到最高在拥有量上的差异程度特别感兴趣。假定我们有一个数量为 N 的总体,每个个体 i 拥有财富 W_i,然后考察一个对总体按财富进行排序而得到的图。我们可以计算出总体中的每个百分之一所拥有的总财富百分比,借此使分布的展示标准化。然后我们画出总体的百分位对应于累积财富的百分比的图。由此得到的图就是总体中财富的洛伦兹分布。不同社会的洛伦兹曲线有着不同形状;总体而言,财富不平等将更大导致图向东南方向延展更远。从这种将总体按资源拥有量排序进行组织

的技术中可以得到几种关于不平等的测量。基尼系数【Gini Coefficent】和财产拥有的上五分位数和下五分位数之比率,是通常在比较经济发展时用到的不平等测量方法。

——Daniel Little

（高勇译校）

参考文献

Rae, D. W., & Yates, D. (1981). *Equalities*. Cambridge, MA: Harvard University Press.

Sen, A. (1973). *On economic inequality*. New York: Norton.

不平等过程(Inequality Process)

不平等过程是一个对于收入不平等或财富不平等的产生过程的模型。它源自经济人类学中社会分层的剩余理论。在此理论中,剩余是指除去生产成本之后的财富,成功竞争到剩余将会使得剩余得以集中,这解释了为什么在考古挖掘中不平等的显著证据首先出现在采集—狩猎社会获取到了农业剩余的地层中。格哈德·连斯基(Lenski,1966)将剩余理论从解释农业社会初期不平等的增加,扩展到了工业革命后不平等的增加中。连斯基假定受教育更多的人能够保有更多的剩余。

不平等过程可以如下表述:“对人们进行随机配对。每个人都去竞争以获取对方的财富,获取概率为50%。每个人有两个特征,财富以及失败后损失的财富比例 ω。赢者得到输者损失的部分。不断重复此过程。”

虽然赢者可以得到收益,但是连续赢的次数是很少的。财富将会流向那些 ω 较小的人手中。在连斯基的假设下,$(1-\omega_i)$ 是教育水平为 i 的工人在不平等过程中的操作化【Operationalization】。

不平等过程竞争的方程解表明,其均衡分布【Distribution】将会近似于一个双参数的伽马概率密度函数【Probability Density Function】:

$$f_{it}(x) \equiv \frac{\lambda_{it}^{\alpha_i}}{\Gamma\alpha_i} x^{(\alpha_i-1)} e^{-(\lambda_{it}x)} \quad (1)$$

其中:

$\alpha_i \equiv$ 有着 ω_i 的个体的形状参数;

$$\alpha_i \approx \frac{1-\omega_i}{\omega_i}$$

$$\omega_i \approx \frac{1}{1+\alpha_i} \quad (2)$$

$\lambda_{it} \equiv$ 有着 ω_i 的个体的形状参数;

$\lambda_{it} \approx (1-\omega_i)$

$$\frac{\left(\frac{w_{1t}}{\omega_1}+\frac{w_{2t}}{\omega_2}+\cdots+\frac{w_{it}}{\omega_i}+\cdots+\frac{w_{It}}{\omega_I}\right)}{\bar{x}_t} \quad (3)$$

$x \equiv$ 财富,被认为是一个大于0的随机变量;

$\bar{x}_t \equiv$ 在时间点 t 时财富的非条件均值;

$w_{it} \equiv$ 在时间点 t 时有着 ω_i 的人口比例。

不平等过程产生了(a)受教育较高者有着更高的平均收入,(b)他们收入有着较低的基尼系数,(c)收入分布的形状取决于教育水平。$f_{it}(x)$ 中 x 的均值为:

$$\bar{x}_{it} = \frac{\alpha_i}{\lambda_{it}} \approx \frac{\bar{x}_t}{\omega_i\left(\frac{w_{1t}}{\omega_1}+\cdots+\frac{w_{It}}{\omega_I}\right)} \quad (4)$$

ω_i 越小,x 的均值越大,教育分布给定的条件下,$w_{it}s$ 是一个常数。伽马概率密度

函数的基尼系数的麦克唐纳和詹森(McDonald and Jensen)表达式为:

$$f_{it}(x)=\frac{\Gamma\left(\alpha_i+\frac{1}{2}\right)}{\sqrt{\pi}\,\Gamma(\alpha_i+1)} \qquad (5)$$

公式(2)给定的条件下,图1显示了$f_{it}(x)$中x的基尼系数将会随着ω_i的增大而上升。从经验来看,教育程度较低者的收入基尼系数将会更高。

图2显示了1987年拟合美国非大都市工人分教育水平的收入分布的$f_{it}(x)$图。图2中伽玛概率密度函数拟合分布的形状参数α_i随教育程度而提高,所以ω_i随教育程度而降低。相同的形状序列可以在所有工业化社会中再现。

塞勒姆和芒特(Salem & Mount,1974)倡导使用双参数的伽马概率密度函数来为收入分布建模,但是并不对其形状参数和标度参数进行约束。不平等过程的伽马概率密度函数模型$f_{it}(x)$的参数是有约束的,两个参数都是ω_i的函数。图2显示了不平等过程的$f_{it}(x)$对33年×5种教育程度=165个分布的单年联立拟合情况。在本例中,$f_{it}(x)$在建模的总体中有一个参数要取五个值,因此只需要估计五个参数值。塞勒姆和芒特的方法需要估计165 × 2 = 330个参数。尽管自由度【Degrees of Freedom】更少,不平等过程的$f_{it}(x)$的拟合效果却几乎一样。克莱伯和科茨(Kleiber & Kotz,2003)在其章节“伽马类型分布”中讨论了不平等过程。

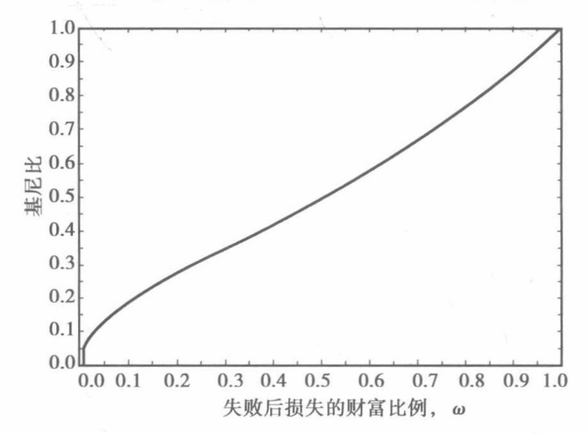


图1 基尼集中比率与ω的关系图
基尼集中比率由形状参数α为$(1-\omega)/\omega$的伽马概率密度函数计算得出。

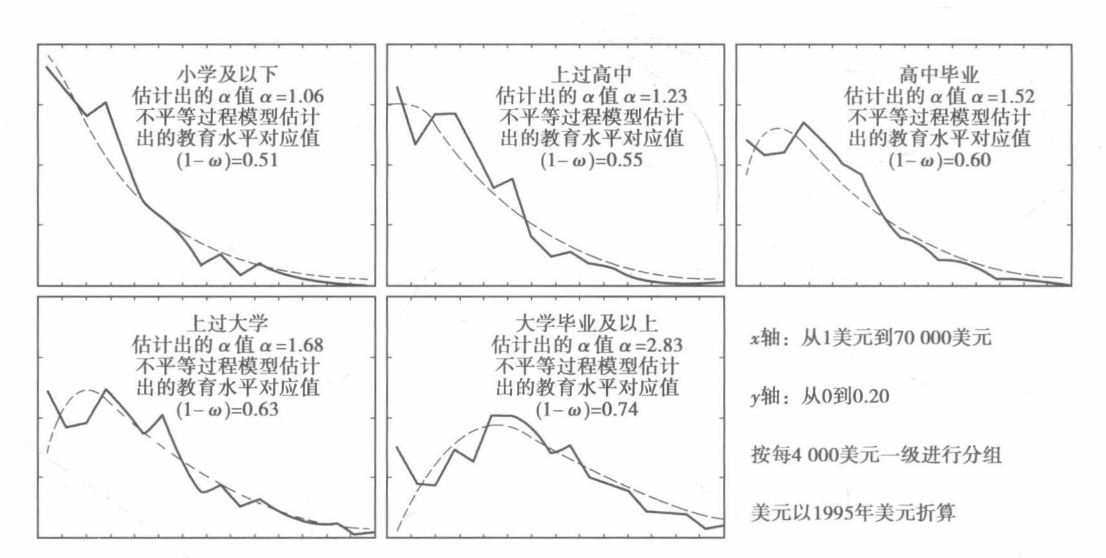


图2 1987年工人的年工资和薪金收入的相对频率分布(美国非大都市居民,图中为实线)及其拟合的不平等过程模型(图中为虚线)。工人年龄在25到65岁。从1965到1995年的数据中进行单年联立拟合得出。

资料来源:三月当期人口调查

——John Angle
（高勇译校）

参考文献

Angle, J. (1986). The Surplus Theory of Social Stratification and the size distribution of personal wealth. *Social Forces*, *65*, 293-326.

Angle, J. (2002). The statistical signature of pervasive competition on wage and salary incomes. *Journal of Mathematical Sociology*, *26*, 217-270.

Kleiber, C., & Kotz, S. (2003). *Statistical size distributions in economics and actuarial sciences*. Hoboken, NJ: Wiley.

Lenski, G. (1966). *Power and privilege*. New York: McGraw-Hill.

Salem, A., & Mount, T. (1974). A convenient descriptive model of income distribution: The gamma density. *Econometrica*, *42*, 1115-1127.

推论(Inference)

参见统计推论【Statistical Inference】。

推论统计学(Inferential Statistics)

广义而言,统计学可以分为两种类型:描述统计学和推论统计学。推论统计学是用来得到样本统计量和总体参数【Parameter】之间关系的结论的一系统程序。一般来说,推论统计学解决两类问题。首先,从一个参数 θ 已知的总体中得到给定样本的概率是多少;其次,从样本统计量中是否可能估计出总数参数 θ。

上述两种操作都需要演绎逻辑而非归纳逻辑;因此,没有闭合证明(closed form proof)能够表明我们的推论是正确的(参见古典推理【Classical Inference】)。但是,利用中心极限定理【Central Limit Theorem, CLT】以及使用合适方法来选择我们的观察,进行确定性可以估计的概率推论是可能的。

如果从均值为 μ 的总体中重复抽取样本量为 n 的随机样本,其样本均值将大致服从正态分布,样本均值的均值将近似于 μ。上述结论既可以经验性地显示,也可以通过 CLT 进行逻辑证明。此外,如果总体元素围绕总体均值以方差【Variance】σ^2 分布,则样本均值的抽样方差为 σ^2/n。样本均值的标准差通常被称为标准误【Standard Error】,以区别于单个样本的标准差。

随着样本量 n 增加,样本均值的分布将越来越接近于正态分布。基于对正态分布的描述,我们知道正态曲线下 68%的面积(或数据点)在从均值向左右各扩展一个标准差的范围内。对于正在处理的样本均值分布,我们也可以得出结论,68%的样本均值将在真正的总体均值向左右各扩展一个标准误的范围内。

由于正态分布是渐近【Asymptotic】的,得到的样本均值可能离总体均值无限远。但是,常识告诉我们,虽然这种情况确实可能发生但是可能性极小。因此,我们有了推论统计学的第一个应用。抽取出一个随机样本,计算出相关的样本统计量,如 $\bar{y}$。研究者常常想知道 $\bar{y}$ 是否等同于已知的总体均值 μ。例如,一个社会学家可能想知道一个移民随机样本的平均收入和总人口的平

均收入是否相同。这个社会学家的虚无假设是 $\bar{y}-\mu=0$，备择假设是 $\bar{y}-\mu\neq 0$。$\bar{y}$ 值可以从样本中计算得出，μ 和 σ^2 可以从普查数据中得到。样本均值离总体参数有多少个标准差，可以用下列式子估计出来：

$$z=\frac{\bar{y}-\mu}{\sqrt{\sigma^2/n}}$$

随着 $\bar{y}-\mu$ 增加，$\bar{y}=\mu$ 的实际可能性就会减小，样本来源的总体不同于现在考虑的这个总体的可能性就越大。样本均值距离总体均值两个标准差以上（即 z 值大于 2）的概率只有 5%；样本均值距离总体均值三个标准差以上的概率则不到 0.3%。

推论统计学解决的第二个基本问题用到了上面提到的许多思想，却是以不同的方式来提出问题。例如，一个民意调查公司在选举日对选民进行随机抽样调查，问他们如何投票。调查者的问题是，是否可以基于调查结果估计出总体值 μ。在此一般的方法是围绕样本估计量 $\bar{y}$ 构建置信区间。使用样本方差 s^2 来作为总体方差 σ^2 的估计量，调查者估计总体的标准差为 $\sqrt{s^2/n}$。

调查者决定需要对结果有 95% 的"把握"。因此，对 $\bar{y}$ 构建一个 95% 的置信区间如下：$\bar{y}\pm 1.96$ s.e. $(\bar{y})$ 或者 $\bar{y}\pm 1.96\sqrt{s^2/n}$。调查者现在得出结论，$\mu$ 大致在 $\bar{y}$ 的两个标准误范围内，这一论断 20 次中有 19 次会是正确的，或者说置信水平为 95%。

这两种基本方法可以进一步复杂化来解决许多问题。例如，我们可能想比较两个或多个样本均值，来决定他们在统计学意义上是否相同。这相当于问这些样本是否来自于同一个总体。我们可以根据能得到的信息来稍微修正一下公式。例如，有时总体方差 σ^2 是已知的，有时它是未知的。如果它是未知的，我们通常从样本方差中估计它。类似地，如果考察的是多个样本，其方差是等同的，我们可以合并它们的方差来得到更好或更准确的总体方差估计。

明显地，推论统计学中的一个核心要素是假定总体的抽样分布已知。根据中心极限定理，用简单随机抽样方法抽取出的许多样本，样本分布都会接近于正态分布。但这一假定并非一定成立。例如在小样本中，样本方差 s^2 往往低估了总体方差 σ^2。对于小样本，我们通常使用 t 分布，t 分布的顶部比正态分布更平，尾部比正态分布更厚。t 分布事实上是一组分布。分布的具体形状取决于自由度，自由度又与样本量相关。

在其他情况下，我们使用其他分布。例如，对于计数数据，我们通常使用卡方分布。在比较方差时，我们使用费希尔（Fisher）首先定义出的 F 分布。偶尔地，抽样分布的函数形式也可能是未知的。在这些情况下，我们有时会使用非参数统计量或者与分布无关的统计量。社会科学家使用许多这样的程序，包括 Gamma ν，Kolmogrov-Smirnov 检验【Kolmogorov-Smirnov Test】，以及各种肯德尔的 Tau 统计量。

随着计算机使计算变得越来越容易，许多科学家使用刀切法（jackknife）和自助抽样推论法【Bootstrapping】建模技术来估计经验分布。这些程序并不总是理想的，但是可以用来成功地估计出中位值或者四分位距【Interquartile Range】这类统计量的抽样分布。如果样本是通过复杂抽样程序或非随机抽样程序得到的，那么建模技术也是非常有用的。

对于结果的阐释也可以基于完全不同的框架。例如，贝叶斯推论【Bayesian Inference】要基于先前研究或者研究者的理论判断对参数作出先验了解（prior knowledge），在推论过程中要包括这些先验了解。另一方面，似然方法认为参数的可能取值并非固定不变，而是构成了一个分布。因此在这种方法中，样本结果的分布函数就

要根据参数取值的可能分布来进行判断。

——Paul S. Maxim

（高勇译校）

参考文献

Agresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Hacking, I. (1965). *The logic of statistical inference*. Cambridge, UK: Cambridge University Press.

Maxim, P. S. (1999). *Quantitative research methods in the social sciences*. New York: Oxford University Press.

强影响点(Influential Cases)

如果样本中的一个或几个案例对分析结果有着非常大的影响,我们称之为具有强影响的观察(influential observation)。

——Kwok-fai Ting

（高勇译校）

强影响统计量(Influential Statistics)

如果样本中的一个或几个案例对分析结果有着非常大的影响,我们称之为具有强影响的观察(influential observation)。换言之,去除这些观察会改变统计结论中的某些方面。因此,具有强影响的观察的存在会损害对多项资料进行概括的目标。例如,极大或极小的几个观察就会使得均值统计量在描述一个分布的集中趋势上不再有用。除了参数估计,无论样本大小,具有强影响的观察也会影响显著性检验【Significance Tests】和拟合优度测量【Goodness-of-Fit Measures】的有效性。由于存在这些潜在问题,很多强影响统计量被构建出来以评估每个观察值在分析中的影响程度。

强影响统计量首先是在线性回归分析中被加以讨论,之后又被扩展到 Logistic 回归【Logistic Regression】分析中。这些统计量试图去测量两个主要的组成部分:杠杆作用(leverage)和残差。杠杆作用是一个观察在迫使拟合函数靠近其观察值上所具有的效力。在回归分析中,那些远离中心(即所有预测变量的均值点)的观察会有更大的杠杆作用,使结果靠近自己。因此,杠杆作用较高的观察是那些有着与众不同的异常特性的观察。残差是分析中观察值与拟合值的差值。残差较大表明有异常存在,因为它们与统计模型反映的期望模式有所不同。之所以对此要加以关注,是因为有着较大残差的观察常常会对回归分析的结果有很大影响;即去除这些观察将会使拟合函数更靠近其余观察。

要评估异常值的影响,画图是一种很有用的可视工具。图 1 显示了对假想数据进行双变量分析时,三个可能对拟合的回归函数有较大影响的观察。观察 1 有着较高的杠杆作用,因为它远离 X 的中心,它会对回归线的斜率产生较大影响。观察 2 靠近 X 的中心,但是它远远地偏离了拟合回归线的期望值。残差较大可能会影响到拟合优度测量和显著性检验统计量。观察 3 既有着较高的杠杆作用,也具有较大的残差,它更可能扭曲分析结果。对于多变量分析而言,

偏回归图更适合用于检验强影响观察,因为它考虑到了模型中的多个预测变量。

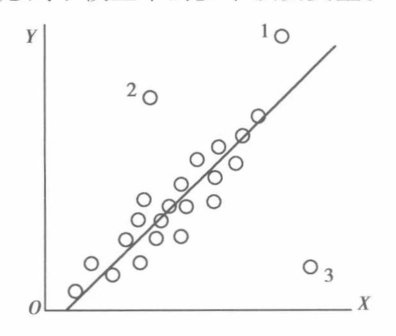


图 1 双变量散点图,其中的回归线是按不存在异常观察时拟合的

对于定量测量来说,帽子矩阵中的值(hat-value)是评估一个观察在影响估计量中的杠杆作用的常用指标。残差通常用学生化残差(studentized residuals)来测量,即残差与其估计标准误差之比。有几个广泛使用的强影响统计量在计算中将杠杆作用和残差综合起来。这些测量的共同思路是,依次去掉每个观察,观察其大致后果,来评估它在分析中的影响。DFFITS 考察在估计过程中包括某个观察与排除某一个观察时,这个观察的两个拟合响应值之间的差异。Cook 距离统计量(Cook's distance)考虑的是排除一个观察对所有拟合值的影响。DFBETAS 统计的是排除每个观察对回应系数的影响。

强影响统计量是有启发性的工具,它提醒分析者注意分析中存在问题的那些观察。虽然有人提出了这些统计量的一些边界值(cutoff),但是这些边界值都只是建议性的,并不是从统计学理论中推导出来的,对此应加以灵活应用。对这些统计量画图也是很有用的方法,可以发现这些存在问题的观察和其他观察之间的主要差异之处。实践中,分析者需要排除这些存在问题的观察之后再重复进行分析,以便充分地评估其实际影响。即使这些观察已经被确认对某一种分析是有较强影响的,这也并不意味着这些观察在最终分析中应当被删除。更重要的是,在寻求解决方案之前,先去探究清楚他们出现异常模型的原因。通常的原因包括编码错误、模型设定有误、纯粹的抽样概率。

——Kwok-fai Ting

(高勇译校)

参考文献

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. New York: Wiley.

Bollen, K. A., & Jackman, R. W. (1990). Regression diagnostics: An expository treatment of outliers and influential cases. In J. Fox & J. S. Long (Ed.), *Modern methods of data analysis* (pp. 257-291). Newbury Park, CA: Sage.

Collett, D. (1991). *Modelling binary data*. New York: Chapman and Hall.

Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. New York: Chapman and Hall.

Fox, J. (1991). *Regression diagnostics*. Newbury Park, CA: Sage.

知情人访谈(Informant Interviewing)

知情人访谈是在田野场景(field setting)下对经过挑选的一小部分知情人进行的深度较深、结构性较弱的(半结构化或非结构化的)访谈。这种访谈区别于其他的可能形式(如认知任务、抽样调查访谈)的特点在于访谈格式的深度、广度、结构与目

标。知情人访谈可以在研究项目的不同阶段使用,以达到各种不同的研究访谈目标。约翰逊和韦勒(Johnson and Weller,2002)区分了自上而下的访谈(top-down interviews)和自下而上的访谈(bottom-up interviews)。自下而上的知情人访谈有助于澄清那些没有被充分理解的论题和知识领域,有助于在研究的随后阶段中构建出更为有效的结构性访谈【Structured Interview】格式(如结构性问题)。自上而下的知情人访谈有助于对现有的结构性问题(如抽样调查【Survey】问题)进行确证和更改,从而可以使现有的访谈格式(如认知任务)更加适用于当下的研究背景。

知情人访谈的一个重要应用是探索一个还没有被充分理解的论题。研究者根据其知识、经验、对指定问题领域的理解来选择知情人(Johnson,1990)。在这些情况下,访谈者扮演了学生的角色,引导访谈以便从这一领域中专家的知情人那里学习。斯普拉德利(Spradley,1979)的《民族志访谈》(*The Ethnographic Interview*)对这种情况下的访谈进行了极其精彩的讨论,他区分了在访谈初始阶段的"泛泛游览(grand tour)"和"深度游览(mini tour)"问题。所谓"泛泛游览",就是向知情人提问,请他向研究者介绍某个地点、场景或某个议题,或请他描述通常如何安排自己的一天,如何与别人交往和互动。深度游览则要就这些一般性问题寻求更为具体的信息。这些非结构性问题的目标是更加了解知情人对社会环境、情景或者知识领域的主观理解。

在这种游览性质的提问中得到的对当地术语和当地议题的理解,有助于向知情人提出更为具体的问题。有许多种更有针对性、更具结构性的知情人访谈格式,如分类学方式(taxonomic approaches)或者自由式回忆启发法(free recall elicitation),通过这些方法都可以从知情人那里得到对指定问题的更有条理、更系统的理解。此外,这些访谈可以用来发展出更为系统的问卷和量表类型(即自下而上的方式),借此可以在知情人(或被访者)中对其信仰、知识等进行比较分析。总而言之,知情人访谈是许多研究形式的重要组成部分,特别是在民族志【Ethnography】研究中,因为它们收集到的资料对研究项目的整体效度【Validity】贡献极大。

——Jeffrey C. Johnson

(高勇译校)

参考文献

Johnson, J. C. (1990). *Selecting ethnographic informants*. Newbury Park, CA: Sage.

Johnson, J. C., & Weller, S. C. (2002). Elicitation techniques for interviewing. In J. F. Gubrium & J. A. Holstein (Eds.), *The handbook of interview research*. Thousand Oaks, CA: Sage.

Spradley, J. P. (1979). *The ethnographic interview*. New York: Holt, Rinehart and Winston.

知情同意(Informed Consent)

自愿的知情同意指的是一个沟通过程,在这个过程里可能的研究对象决定是否参与计划进行的研究。"自愿(voluntary)"意味着没有威胁和过度诱导。"知情(informed)"意味着研究对象在同意之前知道一个理智的人在这些情况下所想知道的一切。"同意(consent)"意味着明确地表明同意参与。实现上述定义的知情同意可能

极具挑战性，特别是当参与者出于情境原因、智识原因或文化原因而人身自主受到损害时，或者不接受那些通常与社会研究和行动研究相伴随的社会规范时。

历史背景

在第二次世界大战后，人们发现纳粹科学家使用囚徒来进行残忍的医学实验。在针对纳粹战犯的纽伦堡审判中，这些罪行也被加以调查，并促成了纽伦堡规范（Nuremberg Code）的产生，规范要求科学家必须得到参与者的知情同意。1974 年，美国国家研究法案（U. S National Research Act）的通过促成了保护生物医学和行为研究中人类受试者全国委员会（National Commission for the Protection of Human Subjects in Biomedical and Behavioral Research，以下简称“全国委员会”）的建立，并且使用机构伦理审查委员会（Institutional Review Boards，IRBs）来监督联邦资助的涉及人的研究（Human Research）。全国委员会对涉及人的研究中的伦理问题进行听证会，去监督知情同意和研究对象的福利被其他关切所遮蔽影响的生物医学研究。在《贝尔蒙报告》（*Belmont Report*）中，全国委员会规定了治理涉及人的研究的三项原则：尊重研究对象的自主权和福利，善行（beneficence）、公正［参看伦理原则（Ethical Principles）］。知情同意是给予研究对象尊重的重要过程。

知情同意的要素

美国针对涉及人的研究的联邦法规（U.S. Federal Regulations of Human Research，以下简称“联邦法规”）制定了知情同意的要素，如下：

（1）试验具有研究性质的声明，对研究目的和预期参与项目的持续时间的解释，对遵循的程序步骤的描述，对任何实验性程序步骤的澄清。

（2）对带给研究对象的任何可理性预见的风险或者不适的描述。

（3）对带给研究对象或其他人的任何可理性期望的收益的描述。

（4）说明对研究对象可能有利的其他治疗方法，如果这些方法存在的话（主要适用于临床生物医学研究）。

（5）如果有保密问题存在，要声明确保保密以及保密的程度。

（6）如果研究中存在风险不是极小，要声明如果发生伤害是否会提供补偿或救治。

（7）说明可以联系谁解答有关研究和被试者权利的相关问题；与研究相关的伤害发生时，可以联系谁去解决。

（8）需要声明参与是自愿的，拒绝参加不会受到惩罚，研究对象可以在任何时间退出且不会受到惩罚。

根据联邦法规，如果研究非如此即不能实施，机构伦理审查委员会可以放弃上述知情同意要素中的某一些甚至全部。但是，在放弃或改变上述要求时，机构伦理审查委员会必须证明研究对研究对象的风险极小；放弃这些要求不会对研究对象的权利和福利有不利影响；如果合适，在参与后会提供给研究对象相关信息。这些要素通常以相当标准化的格式用书面形式写出，并且要有研究参与者的签字，以表明他们已经阅读、理解、同意上述条款。如果研究对研究对象的风险极小，并且研究程序在研究之外的一般情形中并不需要书面同意时，上述信息也可以不是书面格式，同意也不必以签字形式确认。如果泄密会伤害研究对象，并且签字同意是把研究对象和研究联系起来的唯一记录时，出于保密原因也可以放弃书面签字的要求。

知情同意的误用

许多调查者和机构伦理审查委员会似乎认为一份签字的知情同意书就能够实现知情同意的伦理要求。有些同意书又长又复杂,不符合潜在研究对象的文化习惯,几乎大多数人都不能够理解其内容。将知情同意和签字的知情同意书两者混淆,这违背了知情同意的伦理意图,其意图原本是要进行清晰而有敬意的沟通,促进理解和良好的决策,确保参与是自愿的。

研究者和机构伦理审查委员会为什么会采用糟糕的同意程序?针对以人为对象的研究的联邦法规(45 CFR 4)是用来规范美国卫生及公共服务部(Department of Health and Human Services,HHS)资助的研究的。多数 HHS 资助的研究都是生物医学的,因此毫不奇怪,处理以人为对象的研究的联邦法规是以生物医学为模板起草的。这种生物医学的关注点一直给社会研究和行为研究带来很多问题。由于人类(研究)受试保护办公室(Office for Human Research Protections,OHRP)暂停了整个机构的资助研究以应对关于生物医学研究的抱怨,并且将带给研究对象的伤害归罪于机构伦理审查委员会,问题现在变得更严重了。责任的确认方式是检查出问题的机构中的机构伦理审查委员会的文档,发现不完整的文档业务。这样一种大张旗鼓、费时耗力的对研究机构的制裁导致了一种恐惧担心的环境,机构伦理审查委员会认识到采用极度科层化的程序符合他们和机构的利益,因此更加强调文档工作而不是有效的沟通。这种新的管制氛围的结果就是,协议书在批准前要经过多次反复的官僚程序,研究对象必须在又长又复杂的知情同意书上签字。

由于 45 CFR 46 的 A 部分已经被包含到 17 个联邦机构的管制结构中,上述问题变得更为复杂了(这 17 个联邦机构包括住房与城市发展部、司法部、运输部、退伍军人事务部、消费品安全部、环境保护部、国际开发署、航空航天管理局、国家科学基金、农业部、商务部、国防部、教育部、能源部、卫生与公共服务部、社会保障部、中央情报局)。A 部分现在被称为"基本规则(common rule)",管辖所有接受联邦资金的机构中的社会研究和行为研究。结果大多数机构伦理审查委员会就努力地去"照章办事",要求研究对象在同意书上签字,即使这样做极端无礼(如人们不识字时或者处于人们并不信任签字文件的文化中时)。辛格(Singer,1978)和特赖斯(Trice,1987)已经发现,在西方主流文化中,如果要求研究对象在同意书上签字,相当一部分人会拒绝参与调查、研究或实验,但是如果不需要签字,他们会非常乐意参与。在那些愿意签字的研究对象中,大多数人也根本没有阅读同意书。

如果研究者和机构伦理审查委员会记得他们有权对某些特定类型的研究破例,有权进行快速审查,有权根据 45 CFR 46.117(c)放弃签署书面同意书的要求,或者他们承认其他证明知情同意的方式(例如通过他人见证,通过研究者证明进行过同意,通过录音或录像来记录同意过程),前述问题就可以得到避免。最重要的是联邦法规中46.117(c)的第 2 段:"如果研究对于研究对象的风险极小,并且研究程序在研究之外的一般情形中并不需要书面同意时",机构伦理审查委员会可以不要求有签字的同意书。多数社会研究和行为研究带来的风险都极小。"研究程序在研究之外的一般情形中并不需要书面同意"的表述指的是医学程序。就其病史对患者进行访谈,或者让患者填答与其病史有关的问卷,并不需要让他们签署贝尔蒙报告同意书。因此,我们完全可以对调查研究的被访者提类似的个人问

题,而不需要根据“基本规则”让他们签署知情同意书。以下建议可以使我们对“基本规则”进行明智的阐释:

知情同意应当采用友好且易懂的沟通形式。它写出来应当简明扼要、用词平易,即使是略懂读写的研究对象也能够理解。研究对象应当有一个舒适的环境来思考他们被告知的内容,来提出想到的问题。同意过程可能采用多种形式。例如,如果是在一个退休社区中进行老年人的居住情况调查,访谈者应当牢记老年人关心的安全问题,应当牢记他们存在的记忆、听力和视力障碍。开始时,访谈者可以在社区会议中心举行会议,解释访谈的目的,回答大家的问题。他们可以与社区管理层接洽,在社区的内部通信中对讨论内容进行概述,同时附上访谈者的相片。在访谈准备中,他们可以通过电话与每个家庭进行初次接触。在入户时,他们可以带上证明材料以便居民能够识别他们确实是访谈者。访谈者可以简短地口头介绍一下自己的目的、保密措施以及其他需要同意的相关事宜;然后得到被访者的同意;可以酌情提供给研究对象一份书面的信息摘要以便保存。

如果有可能给研究对象带来风险或者不便,就应当提供给他们足够的信息,以便他们能够判断风险是否在他们可接受的范围之内。一个充分的知情同意过程能够把乐于参与者和希望退出者区分开来。有一些风险或不便,如果让人们选择他们可能会乐于接受,但是他们不想他人将这些风险或不便强加于自身。另一些人可能会认为这些经历对自己来说太过紧张、过于冒险、令人不快。

研究者应当说明研究对象在任何时候都可以自由提问。研究对象可能在真正进入研究过程时,才对其参与产生问题或表达关切。例如,研究对象在被问及私人问题时,才会真正重视有关保密的问题。

如果研究对象以挂断电话或者扔掉邮寄来的问卷的方式拒绝参与,那么知情同意应当尽可能简略。礼貌和专业性可能会要求提及研究者及研究机构的身份,以及研究的性质和目的。但是,如果不涉及任何风险、收益、保密议题,就没有必要提及这些内容以及拒绝参与的权利,此时这些细节是无关紧要的。对不相干的信息进行详尽讲述会干扰沟通过程。担保不存在风险,描述用来确保做到保密的措施,这些无关紧要又令人心烦的举动会降低回答率(参看隐私和保密【Privacy And Confidentiality】)。

研究对象的文化规范和生活方式能够表明如何进行知情同意。用与其文化和情境不相容的方式去对待他人是无礼而又莽撞的。例如,要调查无家可归的注射吸毒者,就可能先要花几个星期和他们一起闲逛和交谈,用他们非正式的沟通方式——小道消息(grapevine)来讨论其关心的内容,回答他们的问题,和他们一起协商进行调查的条件。在对难以确定其年龄的离家出走者和街头吸毒青年进行研究前,也应该进行这样的沟通。这些年轻人可能被认为是不受约束的,或者因为找不到其父母而无法得到其父母的同意。那些事实上的文盲者,或者对拿出文件或要求签字持有戒心者,或者那些来自传统文化者,都应当用他们感到舒适的方式来接近他们。在基于社区的研究中,社区的正式领导或非正式领导是守门者(gatekeepers)。聪明的研究者会咨询他们以了解可能的研究对象的关切和感受,怎样招募研究对象最好,怎样规划和实施知情同意过程。对于这些人群的研究协议中应当有证据表明,研究者熟悉这些人群的文化,并且已经相应地组织知情同意和其他研究程序。

研究者和机构伦理审查委员会应当灵活地考虑实施知情同意的诸多媒介。与常用的条文主义的正式同意书相比,录像、手

册、群体讨论、网站等都可能是与潜在的研究对象沟通的更佳方式。

书面的或签字的同意书会给研究对象带来风险,而且签字的同意书是研究对象和研究之间唯一的联系时,应当考虑放弃知情同意。一种有用的程序是让信得过的同事见证口头表示的同意。如果研究对象是文盲,也可以采用这种有见证人的同意,见证人进行签字即可。

在组织研究中,允许进行研究的上级和被研究的下级都应当表示同意。双方都应当有否决权。上级不能强迫下级参与,也不应当获知有关下级是否同意的信息,或者获知其反映信息。

有时候,人们只有在参与之后才能决定是否同意。例如,在简短的街头拦访中,多数被访者只在参与之后才会想到接受同意程序。在访谈之后,研究者才会向他们解释这是一项什么样的研究,他们才可以表明他们是否允许这些访谈作为研究资料被利用。正常的同意程序可以确保做到这一点。

总而言之,知情同意是心存敬意的沟通过程,而不只是一张同意书。研究者应当在心里将这一定义置于首位,然后就可以根据研究情境来选择合适的内容和形式了。

本词条部分内容的初稿得到了罗伯特·莱文的批评意见,在此致谢。

——Joan E. Sieber

(高勇译校)

参考文献

Fisher, C. B., & Wallace, S. A. (2000). Through the community looking glass: Reevaluating the ethical and policy implications of research on adolescent risk and psychopathology. *Ethics & Behavior*, *10*(2), 99-118.

Krueger, R. A. (1994). *Focus groups: A practical guide for applied research* (2nd ed.). Thousand Oaks, CA: Sage.

Levine, E. K. (1982). Old people are not all alike: Social class, ethnicity/race, and sex are bases for important differences. In J. E. Sieber (Ed.), *The ethics of social research: Surveys and experiments* (pp. 127-144). New York: Springer-Verlag.

Marshall, P. A. (1992). Research ethics in applied anthropology. *IRB: A Review of Human Subjects Research*, *14*(6), 1-5.

Melton, G. B., Levine, R. J., Koocher, G. P., Rosenthal, R., & Thompson, W. C. (1988). Community consultation in socially sensitive research: Lessons from clinical trials on treatments for AIDS. *American Psychologist*, *43*, 573-581.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont Report*. Washington, DC: U.S. Government Printing Office.

Sieber, J. E. (Ed.). (1991). *Sharing social science data: Advantages and challenges*. Newbury Park, CA: Sage.

Sieber, J. E. (1996). Typically unexamined communication processes in research. In B. H. Stanley, J. E. Sieber, & G. B. Melton (Eds.), *Research ethics: A psychological approach*. Lincoln: University of Nebraska Press.

Singer, E. (1978). Informed consent: Consequences for response rate and response quality in social surveys. *American Sociological Review*, *43*, 144-162.

Singer, E., & Frankel, M. R. (1982). Informed consent procedures in telephone interviews. *American Sociological Review*, *47*, 416-427.

Stanton, A. L., Burker, E. J., & Kershaw, D. (1991). Effects of researcher follow-up of distressed subjects: Tradeoff between validity and ethical responsibility? *Ethics & Behavior*, *1*(2), 105-112.

Trice, A. D. (1987). Informed consent: Ⅶ: Biasing of sensitive self-report data by both consent and information. *Journal of Social Behavior and Personality*, *2*, 369-374.

U.S. Federal Regulations of Human Research 45 CFR 46. Retrieved from http://ohrp.osophs.dhhs.gov.

工具变量(Instrumental Variable)

一般最小二乘【Ordinary Least Squares, OLS】的关键假设之一就是扰动项与任何解释变量都不相关。这一假设的基本原理是,我们假定解释变量和扰动项对因变量的效应是彼此分离的。但是,如果它们彼此相关,情况就并非如此。由于我们无法分解它们各自对因变量的效应,估计量是有偏且非一致的。假定我们有简单的线性回归如 $Y=\beta_0+\beta_1X+\varepsilon$,其中 X 与 ε 正向相关。X 上下移动时,ε 也随之移动,方向取决于相关的性质。在此,与 X 变动相关的 Y 变动,将与 Y 和 ε 之间的变动相关。这意味着,我们不可能准确地估计 X 和 ε 的总体系数。可以表明,样本量趋向于无穷时,问题将更糟。

存在三种情况,其中解释变量和扰动项的相关可能发生:(a)如果解释变量测量有误差,使得其成为随机的,这时问题可能发生。这就是所谓的变量中的误差问题。(b)如果存在一个滞后的因变量,因此伴有自相关的扰动项时,问题可能发生。(c)如果存在联立性问题时,问题可能发生。也就是,因变量同时引发自变量,而非反过来。

处理这种问题的一种办法主要是使用工具变量代替那个与扰动项相关的变量。工具变量就是原始变量的代理变量,它与扰动项不相关。在上面的简单线性回归方程中,X 的工具变量就是与 X 相关(最好高度相关)但与 ε 不相关的变量。我们定义这样一个变量为 Z,使得 $X=Z+v$,其中 v 是一个均值为 0 的随机扰动项。如果我们将其代入上面的原始方程中,我们得到 $Y=\beta_0+\beta_1(Z+v)+\varepsilon=\beta_0+\beta_1Z+\beta_1v+\varepsilon=\beta_0+\beta_1Z+\varepsilon^*$,其中 $\varepsilon^*=\beta v+\varepsilon$。原始方程中 X 和 ε 相关,而现在方程中的 Z 和 ε^* 则不相关。这是因为 ε^* 包含 βv,它独立于 Z。当我们无限增加样本时,Z 和 ε_i^* 之间的相关接近于 0。手头有这样一个代理变量 Z,我们仍然不能得到 X 系数的无偏估计,因为代理变量并不等于真实变量。但是,当我们增加样本量时,基于 Z 的估计量将会接近于真实的总体系数。

工具变量方法有一个实际而严重的问题,即如何为解释变量找到一个令人满意的代理变量。如果我们有强有力的理论,它可以指引我们。如果没有,我们不得不依赖于技术程序去发现它。例如,在二阶段最小二乘估计【Two-Stage Least Squares】中,我们基于所有外生变量【Exogenous Variable】进行回归来得到内生变量的预测值,然后用此预测值作为内生变量【Endogenous Variable】的代理值。但是,不能保证这种非理论的程序能够提供好的代理变量。作为一般规则,代理变量必须与原始变量高度相关。否则,估计将会是极不准确的。

——B. Dan Wood

Sung Ho Park

(高勇译校)

参考文献

Greene, W. H. (2003). *Econometric analysis* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Gujarati, D.N. (1995). *Basic econometrics* (3rd ed.). New York: McGraw-Hill.

Kennedy, P. (1998). *A guide to econometrics* (4th ed.). Cambridge: MIT Press.

交互(Interaction)

参见交互效应【Interaction Effect】、统计交互【Statistical Interaction】。

交互效应(Interaction Effect)

统计学中的交互指的是两个或多个自变量【Independent Variable】在影响一个因变量【Dependent Variable】时彼此之间存在相互作用(interplay)。可以用两种方式来理解交互作用。第一种方式是,每一个自变量对于结果的效应是条件效应(即取决于另一个变量或另一些变量的取值)。第二种方式是,两个或多个自变量的"合并效应(combined effect)"不同于个体效应(individual effect)的加总;效应具有非加性【Nonadditive】。变量间的交互效应被称为高阶【Higher-Order】效应,每个变量独自的效应被称为一阶【First-Order】效应。

例如,人们防晒意识是其居住地区(多雨的西北地区还是干旱的西南地区)和客观的皮肤癌发病风险高低的函数(参看表1)。

表1 居住地区和发病风险之间不存在交互效应——防晒意识的算术平均值水平是居住地区(西北还是西南)和客观的皮肤癌发病风险的函数

		居住地区	
		西北地区	西南地区
皮肤癌发病风险	低	6	8
	高	7	9

在表1中,发病风险和居住地区不存在交互效应。在西北地区,那些发病风险低的人其防晒意识的均值为6;那些发病风险高的人均值为7,由于发病风险导致的防晒意识差异为1。在西南地区,较高的发病风险同样导致防晒意识提高了1个点,从8提高到9。从另一个角度来看,西南地区的居民比西北地区的居民防晒意识高2个点,无论他们的发病风险是低(分别为8和6)还是高(分别是9和7)。发病风险的影响不取决于地区,反之亦然。最终,发病风险(从低到高)使得防晒意识提高了1个点,地区(从西北到西南)使得防晒意识提高了2个点;我们可以预计,那些居住在西北地区的低风险居民的防晒意识要比居住在西南地区的高风险居民高出1+2=3个点,这和我们的观察是吻合的(分别是6和9),发病风险和居住地区的效应是"可加的(additive)"。

在表2中,发病风险和居住地区存在交互效应。同表1中一样,西北地区的低风险者和高风险者在防晒意识上存在1个点的差异[这被称为风险在某一地区的简单效应(simple effect)]。但是,在西南地区,发病风险导致的差异是其四倍(从8点到12点)。从另一个角度来看,对于低风险者而言,居住地区导致的防晒意识差异仅为2个点(6点到8点);对于高风险者而言,差异为5个点(7点到12点)。居住地区的影响取决于发病风险,反之亦然。表1中在西南地区居住且发病风险高的影响为可加的3个点,但是在表2中,在西南地区居住且发病风险高者的防晒意识比在西北地区居住且发病风险低者高出6个点(分别是12和6)。居住地区和发病风险的合并效应是不可加的,这符合交互效应的第二种定义。

表 2　居住地区和发病风险之间存在交互效应——防晒意识的算术平均值水平是居住地区（西北还是西南）和客观的皮肤癌发病风险的函数

		居住地区	
		西北地区	西南地区
皮肤癌发病风险	低	6	8
	高	7	12

表 2 中的交互效应涉及的变量要么本身就是分类的【Categorical】（居住地区），要么是已经进行了人为分类（发病风险），这种情况发生在方差分析【Analysis of Variance】中。两个连续变量【Continuous Variable】也可以交互，一个连续变量和一个分类变量也可以交互，这种情况发生在多元回归分析【Multiple Regression Analysis】中。图 1 显示了居住地区变量（分类）和发病风险（6 度的风险量表）之间的交互作用。西北地区和西南地区的回归线被分别加以显示［被称为简单回归线（simple regression lines）］。在西南地区（实线），发病风险增加时，防晒意识也显著提高。但是，在西北地区（虚线），发病风险对防晒意识根本没有影响。地区是一个调节变量【Moderating Variable】或调节者（moderator），即能够改变其他变量（在此就是发病风险）与结果之间关系的变量。换言之，发病风险的影响取决于居住地区。两条线之间的垂直距离表明了居住地区对防晒意识的影响。如果发病风险很低（在风险量表中为 1），两条线距离很近；防晒意识的地区差异很小。如果发病风险很高（在风险量表中为 6），地区差异就很大。居住地区的影响也取决于发病风险。最后，发病风险高并且居住在西南地区将导致防晒意识非常高，这与我们关于效应不可加的第二个定义一致。

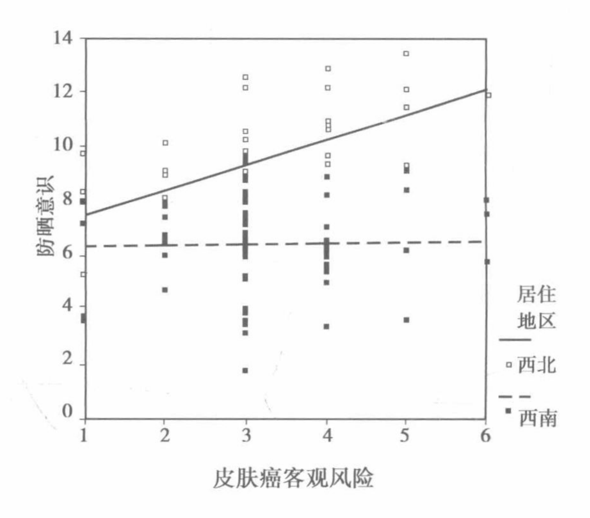


图 1　预测防晒意识中居住地区和皮肤癌客观风险之间的交互

注：居住地区是发病风险和防晒意识之间关系的调节变量。在西北地区，客观风险对防晒意识没有影响；而在西南地区，客观风险对防晒意识有很强的影响。

交互效应的类型

居住地区和发病风险之间的交互效应是协同性的（synergistic）：变量共同作用的效应大于它们独自效应之和。居住在日照强烈的地区（西南地区）并且同时具有高的发病风险导致的防晒意识，要大于它们独自效应的简单加总，这就是“协同性的交互效应（synergistic interaction）”。另一种交互效应被称为缓解性的（buffering），因为一个变量会减弱另一个变量的影响。压力作为一个自变量会增加精神健康受损的风险。从家庭和朋友中获取社会支持会减弱（或缓解）高压力对精神健康的影响，这就是“缓解性的交互效应（buffering interaction）”。

图 1 中显示的效应是定序交互【Ordinal Interaction】；即无论一个变量的取值如何，另一个变量水平的排序保持不变。在这里，无论发病风险取值如何，西南地区居民的防晒意识总是高于西北地区居民，这是定序交

互的特征。交互效应也可能是“逆序的”(disordinal,即展现出了逆序现象【Disordinality】)。比如说,在高风险者中,那些居住在西南地区的人比居住在西北地区的人防晒意识更高。但是,在低风险者中,那些居住在西北地区的人防晒意识更高。交互效应是逆序的。这也被称为“交叉交互(cross-over interaction)”:一个变量的哪个值会更高,这要取决于另一个变量的取值。

历史发展

从1950年代到1970年代,分类变量(即方差分析中的因素)的交互效应的特征和解释是重点所在。科恩(Cohen,1978)将关注重点转向了多元线性回归中连续变量的交互效应。十年之后,雅卡尔、图里西和旺(Jaccard, Turrisi, and Wan,1980)以及艾肯和韦斯特(Aiken and West,1981)发展出了连续变量交互效应的事后探寻(post hoc probing)方法和解释方法。目前的焦点问题是将交互效应的特征界定扩展到回归的其他形式(如logistic回归【Logistic Regression】)当中,以及扩展到结构方程模型【Structural Equation Modeling】当中潜变量间的交互效应。

——Leona S. Aiken

(高勇译校)

参考文献

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Cohen, J. (1978). Partialed products are interactions; partialed powers are curve components. *Psychological Bulletin*, *85*, 858-866.

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression.* Newbury Park, CA: Sage.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

截距(Intercept)

截距的定义为当一条直线或曲线穿过 y 轴时变量 Y 的取值。在二维图中,截距就是当 X 为零时 Y 的取值。截距的单位与 Y 的单位相同。

回归【Regression】分析中对截距的解释有时会有问题:变量 X 的取值范围往往没有涵盖零值,因此 X 为零时 Y 的取值在实际情景或政策情景中就变得没有意义了。

但是,截距是回归方程的重要组成部分:

$$Y = a + bX$$

或者

$$Y_i = \sum_j b_j X_{ij} + \varepsilon_i$$

下标 j 表示不同的变量,从一个常数项 b_0 开始算起。下标 i 表示样本数据集中的个案。在简单二元方程中的截距 a 对应于多元方程中的常数项 b_0。b_0 可以视为第一个变量 X_0 的乘积因子常数(multiplicative constant),而 X_0 的值总是等于1。因此,变量 $X_0 = 1$ 被生成,以便整个方程中包括截距项。

用图可以说明在有些线性方程和非线性方程中,截距项都是不适用的。有些曲线与 x 轴没有交叉,因此就没有截距。但是,对于线性回归来说总是存在 y 截距的,除非

斜率是无限的。在进行估计前可以将 y 截距设定为零。

下面的例子可以说明即使自变量为零在合理取值范围之外,截距仍然是有政策意义的。为了研究许多国家都存在的不同性别间的报酬差异,人们认真研究了男性和女性每小时报酬的回归方程(Monk-Turner & Turner, 2001)。也认真研究了其他报酬差异,如主体民族和少数民族间的报酬差异。使用英国2000年的数据,包括教育最高水平的虚拟变量的工资方程可以表明男性和女性中教育对工资的影响(Walby & Olsen, 2002)。工资的对数被用来代表分布偏斜的工资变量,统计方法用的是线性回归。下面显示的估计是基于一系列教育的虚拟变量,进而产生了一个样条估计(spline estimate)。图中显示的是预测值和正规教育年数之间的关系。其中的曲线是对样本估计的最佳拟合曲线。

用数学术语来说,真正的截距发生于 $x=0$ 时,仿佛教育年数为零是可能的。但是,图1中看到的截距项(男性是5.75英镑,女性是4.25英镑)却构成了不同性别之间存在报酬差异的部分论据:女性曲线的截距比男性曲线的更低。许多争论就是围绕截距差异展开的(Madden, 2000)。

在回归【Regression】分析中,这些方程中的常数项取决于方程中的全部32个自变量(详情请参见 Taylor, 2000; Walby & Olsen, 2002)。但是,回归常数项 B_0 并非在这个特殊的 x-y 图中看到的截距。B_0 是所有 x_i 的多维空间中的常数项。但是,显示某一个 x_i 和 y 的预测值之间关系的简图在 x_i 轴上会有一个看得到的截距。在此必须假定其他 x_i 的隐性取值;它们通常被给定为其均值,这些取值会影响在 x-$\hat{y}$ 图中看到的截距项(即 y 在 x_i 域中的预测值)。截距与斜率可视的两维图与复杂的多元回归【Multiple Regression】的数学运算间存在着复杂的关系。

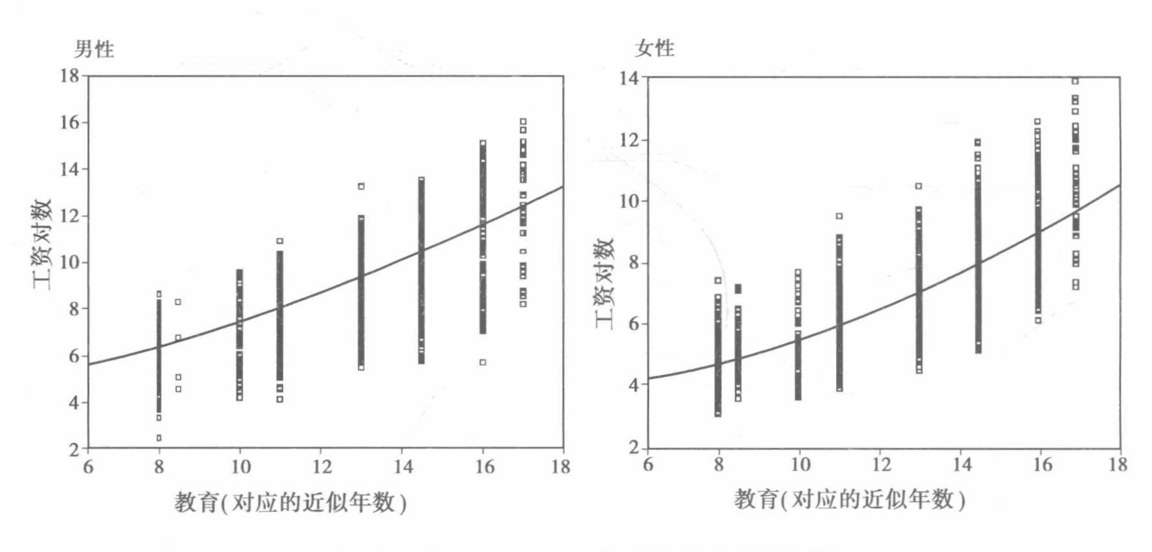


图1 用回归预测的小时工资(教育作为虚拟变量)

数据来源:英国家庭面板调查,第9期,1999—2000(ESRC, 2001; Taylor, 2000)

——Wendy K. Olsen

(高勇译校)

参考文献

Economic and Social Research Council Research Centre on Micro-Social Change. (2001, February 28). *British Household Panel Survey* [computer file] (Study Number 4340). Colchester, UK: The Data Archive [distributor].

Madden, D. (2000). Towards a broader explanation of male-female wage differences. *Applied Economics Letters*, *7*, 765-770.

Monk-Turner, E., & Turner, C. G. (2001). Sex differentials in the South Korean labour market. *Feminist Economics*, 7(1), 63-78.

Taylor, M. F., with Brice, J., Buck, N., & Prentice-Lane, E. (Eds.). (2000). *British Household Panel Survey User Manual Volume B9: Codebook*. Colchester, UK: University of Essex.

Walby, S., & Olsen, W. (2002, November). *The impact of women's position in the labour market on pay and implications for UK productivity*. Cabinet Office Women and Equality Unit. Retrieved from www.womenandequalityunit.gov.uk.

内在信度(Internal Reliability)

社会科学中使用的多数测量(measures)都是基于少量题项(items)构建的,这些题项又是从试图测量这一构念【Construct】的大量可能题项中抽取出来的。如果测量中对各个题项(或者各组题项)的应答(responses)缺乏一致性,就很难评估对测量中某些题项的应答是否可以推广到所有用于测量这一构念的题项总体中。换言之,高度的内在一致性表明对题项的应答之间存在高度的彼此相关性。内在信度提供了对于测量当中各个题项或者各组题项的应答之间的一致性程度的估计。

有许多种内在信度估计量,包括Cronbach α【Cronbach's Alpha】、Kuder Richardson 20(简称KR20)、折半信度【Split-Half Reliability】、分层阿尔法(stratified alpha)、极大信度(maximal reliability)、拉贾系数(Raju coefficient)、克里斯托夫系数(Kristof's coefficient)、安戈夫-费尔特系数(Angoff-Feldt coefficient)、费尔特系数(Feldt coefficient)、费尔特-吉尔默系数(Feldt-Gilmer coefficient)、哥特曼λ_2、最大化λ_4、霍伊特的方差分析(Hoyt's analysis of variance)。在对一组参与者进行某种形式的测量之后,就可以计算出所有上述估计量来。在以下部分中,我们只讨论KR20、折半信度、分层阿尔法和极大信度。

如果测量只包括二分应答选项(例如,是/否、正确/错误等),克龙巴赫的阿尔法就会呈现一种特殊形式的KR20,它可以表示为:

$$\left[\frac{n}{n-1}\right]\left[\frac{\sigma_X^2-\sum p_i(1-p_i)}{\sigma_X^2}\right]$$

其中p_i是在题项i上回答是或正确的人口比例,σ_X^2是在由n个部分组成的测试中得分的方差。

折半信度是一种被广泛使用的信度估计,定义为一个测试的两半之间的相关【Correlation】(r_{12}),然后再用测试的总长度进行修正。根据斯皮尔曼-布朗校正公式(Spearman-Brown Prophecy formula),折半信度可以用公式$\frac{2r_{12}}{1+r_{12}}$计算得出。它也可以将方差【Variance】考虑进来,用如下公式计算:$\frac{4r_{12}\times\sigma_1\times\sigma_2}{\sigma_X^2}$,其中$\sigma_1$和$\sigma_2$是测试的两半的标准差【Standard Deviation】,σ_X^2是整个

测试的方差。当折半信度由后一个公式定义时,克龙巴赫的阿尔法相当于折半信度的平均值。但是,如果折半信度是用斯皮尔曼-布朗校正公式计算的,只有在所有题项方差相等时,克龙巴赫的阿尔法才会等于折半信度的平均值。

如果一个测试 X 不是单一维度的,而是分成了内容有所不同的 i 组,分层阿尔法和极大信度就是比克龙巴赫的阿尔法及其相关估计更好的选择。分层阿尔法可以用如下公式计算:

$$1 - \sum \frac{\sigma_i^2(1 - \alpha_i)}{\sigma_X^2}$$

其中 σ_i^2 是第 i 组的方差,α_i 是第 i 组的 Cronbach α 系数。如果组内的所有题项彼此具有等同性,但是不同组间的题项的信度和方差不同,这时极大信度估计就更为合适(Osburn, 2000)。极大信度可以由下式得到:

$$R_I = \frac{A}{\frac{I}{[1 + (I - 1)]\rho} + A}$$

其中 I=组数,ρ 是组间的共同相关系数,$A = \frac{n_1\rho_1}{1-\rho_1} + \frac{n_2\rho_2}{1-\rho_2} + \cdots + \frac{n_i\rho_i}{1-\rho_i}$($n_i$ 是第 i 组中的题项个数,ρ_i 是第 i 组的信度)。

总体而言,与内部信度或者重测信度【Test-Retest Reliability】相比,复本信度(alternate-forms reliability)是更好的信度估计。遗憾的是,估计复本信度所需要的条件在实践中不易满足(参看信度【Reliability】)。与重测信度相比,内部信度更加灵活,不易受练习、疲劳、记忆、动机等问题的影响。它的灵活性在于它只需要一次实施过程,而且可以适用于多种情景,包括有多个题项的测试、有多个子量表的测试、有多个评价者共同打分的测试。在如下情况下,内部一致性信度通常不是合适的信度估计:(a)测试的内容是异质性的;(b)从两个或多个异质性测试中得到一个综合得分(battery score);(c)测试具有快速完成的特征。

——Peter Y. Chen
Autumn D. Krauss
(高勇译校)

参考文献

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*, 98-104.

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 105-146). New York: Macmillan.

Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods*, *5*, 343-355.

内在效度(Internal Validity)

内在效度指研究者从某一经验研究的结果中进行因果(causal)推论的信心。坎贝尔和斯坦利在一篇很有影响力的论文"实验与准实验研究设计"(Campbell & Stanley, 1963)中,将内在效度视为实验研究的必要条件。这是因为做实验的最终目的就是检验自变量【Independent Variable】和因变量【Dependent Variable】之间的因果关系。因此,评价实验结果的主要标准就是是否可以从其结果中得到有效的因果结论。但是,对于任何试图从中推论出两个变量间因果关系的相关研究发现来说,内在效度的概念也

完全适用。

内在效度并不意味着某一自变量就是因变量存在差异的唯一原因,只是说它具有一些独立的因果作用,澄清这一点很重要。例如,个体对一个社会项目的态度差异可能是由许多因素引起的,包括个人经历、自我利益、意识形态等。但是如果接收了某一种劝服性信息的人群比那些没有接收这一信息的人更赞成这一项目,那么我们感兴趣的问题就是这一信息是否在引发这种差异中起到了作用。只有在其他相关因果因素和个体是否接收这一信息无关的前提下,我们才可以得到这一结论。换言之,此时不存在合理的备择解释来说明态度与信息接收之间的共变。如果有相关的竞争解释因素(rival factors)存在,那么这种情况通常被称为实验处理中存在混杂【Confounding】,因为要研究的变量的潜在效应不可能从这些其他潜在原因变量的效应中分离出来。

在这个例子中,因果推论的最明显挑战是接收信息可能存在自选择。如果研究参与者可以自由选择他们是否接收劝服性信息,那么很有可能那些已经持有赞成态度的人会进入接收信息的组中。在这种情况下,给出信息后两组之间的态度差异可能早已注定,与信息本身毫无关系。这并不一定意味着信息没有因果影响,只是说我们有一个合理的备择解释,这使得这一结论变得不能确定。正是由于这个道理,随机分配【Random Assignment】是优秀的实验设计的重要特征。将参与者随机分配到自变量的不同水平中,这排除了自选择这一对内在效度的威胁。

对内在效度的基本威胁

如果有理由相信即使不操纵自变量的施加,得到的因变量差异也会出现,那么这项研究就没有内在效度。除了自选择这一潜在的混杂因素外,坎贝尔和斯坦利(Campbell & Stanley, 1963)描述了对内在效度的其他种类的威胁。这些因素可能会引起因变量的差异,但是只有在研究实施中这些不相干的变量和我们感兴趣的自变量存在相关时,它们才会成为内在效度的威胁。坎贝尔和斯坦利(Campbell & Stanley, 1963)讨论的潜在混杂因素包括以下类型:

历史。在两个不同时点测量出的因变量结果差异,可能是两次测量时点之间发生的某些事件导致的,而不是由于实验变量导致的。如果因变量是在自变量施加前和施加后测量的,那么历史可能会引发问题,此时其他干预事件可能是变化的真正来源。

发育。在因变量测量的不同时点之间发生的另一种效应,涉及研究中参与者内在条件的变化,如年龄增长、更加疲惫、失去兴趣等(这些因素被称为发育效应,但是这一种类的某些表现,如变得疲惫,通常并不被认为和生理发育相关)。

测试。参与者在第二次接受自变量实施时的表现可能会受到以下事实的影响:他们以前已经接受过一次测量了。因此,测试本身可能成为因变量变化的一个来源。如果研究中的某些个体接受过前测而其他人没有接受过,这也可能引发问题。

仪器。因变量得分在不同时间的变化可能是由于测量"仪器"的性质变化而引起的(如观察者态度的变化、测试打分者越来越草率粗心等),而不是由于被测量的参与者的变化引起的。

统计回归。缺乏信度,或者测量误差,会导致不同时刻下测量分数的变化,如果基于首次测量时的极端得分来选择参与者,那么这些得分就可能导致曲解。

退出实验率。如果进行组间比较,任何

可能导致组间的退出实验比率不同的选择程序或者处理差异，都可能是最终测量中组间差异的真正原因。

选择与历史的交互作用。如果参与者有差别地被选择进入用于比较的组，这些经过特别选择的组可能在历史、发育、测试等方面有所差异，这可能导致因变量在最终测量中的差异。

需要再次强调的是，当且仅当研究中的这些因素与感兴趣的自变量差异相关时，这些因素的存在才会损害内在效度。自变量之外的事件可能会影响到因变量结果，但是如果这些事件与自变量不存在系统相关，它们就不会降低内在效度。

——Marilynn B. Brewer

（高勇译校）

参考文献

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171-246). Chicago: Rand McNally.

Crano, W. D., & Brewer, M. B. (2002). *Principles and methods of social research* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

网络调查（Internet Surveys）

“网络调查”一词是指通过网络收集数据的调查。网络调查对于调查专业的影响是以前各种创新所不能比拟的。就在网络（更具体的是指万维网）被宽泛使用的很短时期里，网络调查软件产品、在线样本组（online panels）、提供网络调查服务的调查研究机构的数量激增，网络调查既被作为一种单独的数据收集模式，也与传统调查研究方法组合使用。但是，业内的接受程度却并不一致，有人热烈欢迎网络调查，并认为它将会取代所有其他调查数据收集方式，另一些人将网络调查视为对调查行业未来的威胁。虽然有关网络调查的价值仍有诸多未知之处，对它们的实用效果的研究正在快速推进。

网络调查的类型

网络在调查数据收集中有几种不同的使用方式。一种区分是在被访者机器（客户端）上执行的调查和在调查机构的互联网服务器（服务器端）上执行的调查。主要的客户端调查方法包括电子邮件调查和可下载的可执行文件调查。在上述每种情形中，量表被传输给样本中的人，然后他可以使用电子邮件软件中的回复功能，或者在文字处理软件中输入回答，或者使用他们电脑中安装的软件，来回答调查问题。完成之后，答案被传输返回给调查机构。服务器端系统通常要求样本中的人与网络连接时通过浏览器完成调查，答案同时被即时地传输给服务器。自动化调查量表的交互特征是用互联网服务器上的脚本完成的。这两种方法的关键区别在于被访者在完成调查时网络连接是否要打开。互联网调查（Web Surveys）是后一种类型的主要例子，到目前为止已经成为现在流行的网络调查的主导形式。

也有一些对上述两种基本类型的变异。例如，一些调查是由单个可流动的 HTML 表

格组成的,填答过程中不需要和服务器进行交互,被访者填答完成之后按下"提交"键来传输信息。与此类似,用JavaScript、DHTML或者其他类似工具也可以在HTML中嵌入客户端的交互功能。这些工具不用互联网服务器也能实现许多动态交互功能。

典型的互联网调查包括一个或多个调查问题,被访者要与万维网连接通过浏览器来回答这些问题。对被访者的抽样可以采用从自选择样本到基于清单的方法的许多形式(有关评论请参看Couper, 2000)。如果是基于清单的抽样方法,通常要通过电子邮件邀请被访者参与调查,邮件信息中包括了调查网址,以及进入调查所需的用户名和密码。互联网调查也越来越多地以混合模式设计进行,邮寄调查被寄送给被访者,但被访者也可以选择在线完成调查。

网络调查的优势

网络调查的首要优势是成本。虽然固定成本(硬件和软件)比邮寄成本更高,但数据收集的单位成本微不足道。通过完全自动化的方法,可以省去邮寄成本和印刷成本,还有数据录入和相关处理成本。这些既减少了数据收集成本,也减少了获取调查结果所需的时间。

互联网调查的另一好处在于它们相对于其他数据收集方式的速度。很多情况下,相当比例的被访者能够在发出邀请的24小时内进行回答。

互联网调查还具有自填带来的优势,包括消除或降低了诸如社会期许性回答之类的访员效应。回答者【Respondents】可以按自己的节奏回答,对过程有更大的控制。同时,互联网调查还有计算机带来的所有优势。这些优势包括可以使用复杂的分岔题目、随机化、编辑检查等,这些特点在计算机辅助访谈【Computer-Assisted Personal Interviewing】中都是常见的。最后,这些调查还可以利用互联网丰富的图形性质,例如量表中可以包括色彩、图形及其他多媒体元素。

网络调查的劣势

网络调查的最大缺点就是代表性问题。网络连接还远未普及,这意味着总体【Population】中有很大比例用这种方法是接触不到的。网络技术覆盖的人和未覆盖的人是有差异的,因此覆盖偏差【Bias】就进入了调查估计中。当然,对于研究者感兴趣的某些总体来说(如学生、高科技行业的雇员、一个网址的访问者等),这并不是大问题。另一个相关的挑战是抽样。网络用户的清单并不存在,对于网络来说,类似于电话调查中随机数字拨号【Random Digit Dialing, RDD】的方法也不可能开发出来,因为电子邮件地址格式的差异太大了。

无回答【Nonresponse】是另一个问题。网络调查中的回答率通常低于对应的纸介方法(参见Fricker & Schonlau, 2002)。在样本中的人可以选择纸张还是互联网填答的混合模式设计中,绝大多数人选择纸介方法。

另一个潜在的缺点是绝对的匿名性。通常没有什么能够保证被寄给邀请的人确实填答了调查问卷,也不能保证他们的填答是认真的。虽然存在这种担心,但从网络调查得到的结果和其他调查数据收集方法相差无几,表明这个问题并不需要太多担心。

小结

虽然网络调查是相对新近的现象,但它已经迅速扩展开来,已经有许多不同的方式来设计和执行在线调查。代表性问题可能限制了网络调查在一般性人群的研究中的

用处,但是它非常适用于对特殊人群(有较高网络使用率的群体)的研究,以及一些代表性并不太重要的研究(如实验研究)。有鉴于此,随着我们不断发现一些关键性挑战的前沿解决方案,不断利用技术提供的机遇,网络调查很可能将继续对调查行业产生影响。

——Mick P. Couper

(高勇译校)

参考文献

Couper, M. P. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, *64*(4), 464-494.

Fricker, R. D., & Schonlau, M. (2002). Advantages and disadvantages of Internet research surveys: Evidence from the literature. *Field Methods*, *14*(4), 347-365.

内插法(Interpolation)

内插法指估计在两个已知取值之间的函数【Function】值或序列值。不同于从已知范围的取值中估计这一范围之外取值的外推法【Extrapolation】,内插法是要从已知范围的取值中估计这同一范围之中的取值。

广义而言,社会学中用到的内插法有两种。一种是数学内插法,另一种是统计内插法。在数学内插法中,我们通过假定两个已知数据点间存在确定性关系来估计未知值。例如,假定我们想找到自由度为 11.3 的卡方【Chi-Square】变量的 99% 临界值(例如,参见 Greene, 2000, p.94)。注意,卡方表中只给出了整数取值的情况。但是事实上,卡方分布的定义对于非整数取值也是成立的。在本例中,我们可以按如下方法用内插法得到非整数取值的结果。首先,从表中得到自由度分别为 11 和 12 时的 99% 卡方分布临界值。然后,在自由度的倒数间进行线性内插。这样,给定 $\chi^2_{11} \leqslant 0.99 = 24.725$ 和 $\chi^2_{12} \leqslant 0.99 = 26.217$,我们有:

$$\chi^2_{11.3} \leqslant 0.99 = 26.217 + \frac{\frac{1}{11.3} - \frac{1}{12}}{\frac{1}{11} - \frac{1}{12}} \times (24.725 - 26.217) = 25.209$$

本例中使用的是线性内插法。但是,在其他问题中,内插法可以用非线性函数来插入未知取值,如余弦函数、多项函数、三次样条函数或其他函数。所选择的数学内插方法应当与希望得到的结果一致。

统计内插法依赖全部数据集的动态或分布【Distribution】的信息来估计位于已知取值之间的未知值。在此要注意它与数学内插法的差异。在前面的例子中,我们只是在两个数据点之间画了一条直线或曲线,而没有考虑数据集的具体内部结构。但是,在统计方法中,我们要考虑内在于数据集的具体动态或分布,并利用这些信息得到未知值。我们可以通过发现其他数据集,或者就是通过对数据的内在结构进行假定,来获取这些信息。例如,假定我们有一个变量 X,它是季度时间序列【Time-Series】数据,我们想将它转换成月度时间序列数据。一种办法是找到一个月度时间序列数据 Y,它与原始的 X 序列高度相关。我们首先去除两个数据序列在季度波动上的差异,然后将 Y 的月度序列投射回 X 上来。得到的结果可能有所差异,这取决于我们如何去除 X 与 Y 之间的季度差异和如何将 Y 投射回 X 序列上来。但是,假定我们找不到与 X 相关的数据序列,我们也可以简单地假定一种内在动

态。例如,我们可能有某个日期之前的季度数据,以及这个日期之后的月度数据(如1978年之前和之后的密歇根大学消费者情绪指数)。我们可以简单地假定,月度数据结构与先前季度数据中未知的月度动态是一致的。利用上述假定,我们可以构建一个 *X* 序列的 ARIMA 模型,其中设定一种自回归【Autoregression】、移动平均【Moving Average】和随机游走【Random Walk】的组合,然后再把这个设定投射到 *X* 的未知月度序列上。

有许多种技术上前沿的内插方法,都是从前面讨论的这两大类别中衍生出来的。但是,需要强调的是,没有一种方法能够消除真值不存在时,内插法中内在的不确定性。这是因为根本不可能发现未知数据的真值。在有些情况下,内插法甚至可能对真实的数据结构造成严重的扭曲。因此,在生成和使用内插数据时要多加小心。

——B. Dan Wood
Sung Ho Park
(高勇译校)

参考文献

Friedman, M. (1962). The interpolation of time series by related series. *Journal of the American Statistical Association*, *57*, 729-757.

Greene, W. H. (2000). *Econometric analysis* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

诠释套路(Interpretative Repertoire)

社会心理的话语分析【Discourse Analysis】部分地旨在确定讲话者在描述事件、选择立场、进行论辩时所用到的解释性资源(explanatory resources)。诠释套路是用于确定与某一特定主题或话题相关的此类资源的一个术语。通常,人们在具有意识形态重要性的访谈资料中寻求这些主题或话题,如种族、性别和权力关系。在致力于勾连参与者的描述实践细节与实践者立足自身的更为宽广的意识形态和历史形成的分析中,诠释套路占有一席之地。

诠释套路的本质在于,描述任何特定事件形态都存在不止一种套路。通常会有两种相互替代或相互冲突的套路,但也可能有更多种。套路是反复出现的词语形象、隐喻、修辞格(figures of speech)、解释方式。要找到这些套路不存在精确定义或便捷之道。找到这些套路的程序是要收集与话题相关的话语资料,考察它们以发现反复出现的诠释模式。这类似于发展扎根理论【Grounded Theory】的阶段,但诠释套路分析的特点在于,它与话语资料的细节紧密结合在一起凸现出话语分析的程序,而不是得到一系列抽象的编码和类型。诠释套路最好用示例来进行说明。

奈杰尔·吉尔伯特和米歇尔·马尔基(Gilbert & Mulkay,1984)在其对科学话语的研究中最早使用了这一术语,他们在研究中发现了两种彼此相反、在功能上相异的解释类型。存在一种经验主义套路(empiricist repertoire)——对结论和理论选择进行的不带个人色彩的、基于方法的、资料驱动的叙述,通常在实验报告中出现。也存在一种机动主义套路(contingent repertoire)——取决于个人动机、洞见和偏差;取决于社会情境;取决于信奉——这是一个猜想和灵感有用武之地的领域,是结论和理论选择引发了支持它的经验工作,而不是经验工作导出了结论和理论选择。在主张事实和支持心仪的理论时使用经验主义套路,在解释为什么以及何时事情出错时使用机动主义套路,特别是在解释对方实验室为什么出错时,解释不

可信的发现为什么出现时。

玛格丽特·韦瑟雷尔和乔纳森·波特(Wetherell & Potter,1992)就种族和民族问题对新西兰人进行了访谈。他们在有关文化和文化差异的谈话中发现了一种遗产套路(heritage repertoire)。这包括诸多描述、隐喻、修辞格和论据,它们围绕保持传统、价值、仪式的重要性而聚合在一起。另一种治疗套路(therapy repertoire)将文化定义为疏远的毛利人为了再次团结和健康而需要的东西。每种套路不是不同的讲述者用来表达不同的态度或信仰的,而是同一个人就会使用两种套路,因为每种套路有其特殊的修辞用法。例如,治疗套路在谈论犯罪和教育失败时就是有用的。此类套路虽然常常是彼此矛盾的,但是通常参与者每次使用时将其视为没有任何问题的常识。诠释套路的这种常识地位使得它与意识形态研究联系起来。

——Derek Edwards

(高勇译校)

参考文献

Gilbert, G. N., & Mulkay, M. (1984). *Opening Pandora's box: A sociological analysis of scientists' discourse*. Cambridge, UK: Cambridge University Press.

Wetherell, M., & Potter, J. (1992). *Mapping the language of racism: Discourse and the legitimation of exploitation*. Brighton, UK: Harvester Wheatsheaf.

诠释性传记(Interpretive Biography)

诠释性传记方法是指在研究中使用和收集描述个体生活转折点时刻的个人生活文件、故事、讲述、叙事(参见 Denzin, 1989, Chap. 2)。传记方法的主题(subject matter)是人的生活经历。如果是用第一人称写的,它被称为自传、生活故事或生活史;如果是他人观察其生活而写的,它被称为传记。

如下假定和论据构建了传记方法的使用。生活只能通过其回溯推理(representation)来了解,包括诸如故事、个人叙述或参与仪式行动等表演(performance)。生命体验【Lived Experience】及其回溯推理正是社会学的主题。社会学家必须学会如何勾连和联结在传记中充满意义的经历和当下的社会,以及更大的后现代时期的形塑文化和意义的制度。这些经历的意义正是在经历它们的那些人的表演中被赋予。对方法的执着,对效度【Validity】、信度【Reliability】、概推性【Generalizability】的执着,对传记方法的理论相关性的执着,必须被搁置一旁,以便更有利于关切意义、活动和诠释。学传记方法的学生必须学会如何使用表演理论、文献诠释、批判主义等策略和技巧。他们必须将方法的使用与批判理论在最近的结构主义【Structuralist】和后结构主义【Poststructuralist】发展一致,这些批判理论关心的是社会文本的阅读和写作。这也会涉及诠释学【Hermeneutics】、符号学【Semiotics】、批判种族理论【Critical Race Theory】、女性主义【Feminist】和后殖民理论,以及教学文化研究。

生活和构建生活的传记方法都是文学创作的产物。生活是一种武断的建构,受到当时文化写作实践的约束。这些文化实践导致了性别化的知情他人(gendered, knowing others)的出现和影响,这些他人将对象置于熟知的社会空间当中,其中生活有了开端、转折点和清晰界定的结束。这些文本创造了"真实"的人,对其进行了推测性的事实陈述。事实上,这些文本是叙述性的

虚构作品,是从他们讲述的生活这一块布料上裁剪下来的。一个作者在写作传记时,他或她把自己也写进了写作对象的生活之中。读者阅读传记文本时,文本是在通过读者的生活被阅读的。因此,作者和读者共谋创造了他们写作和阅读到的生活。接下来,生成的文本会围绕着写到的那个"真实"的人的生活踪迹而聚集起来。

这些假定和立场取决于如何定位和诠释传记材料中的记述对象这一问题,并因此问题而得以构建。所有的传记研究都会假定一个已成过往的生活(a life that has been lived),一个能够被研究、构建、再构建和书写的生活。在当前语境下,生活指涉两种现象:生命体验或有意识的存在和人。一个人既是一个自我意识的存在,也是一个被命名的文化主体或文化创造。人的意识同时既指向思想和体验的内在世界,也指向事件和经历的外在世界。内在世界和外在世界,这两个世界是同一过程的正反两面,因为内在体验和外在经历之间并不存在明确的分界线。传记方法认识到了人的存在这一事实,因为其突出特点就在于,在个人文档这种述行性文本(performative text)中这两种结构被连接与记录下来了。

——Norman K. Denzin

(高勇译校)

参考文献

Denzin, N. K. (1989). *Interpretive biography*. Newbury Park, CA: Sage.

诠释互动论(Interpretive Interactionism)

诠释互动论研究在彼此互动的个体生活中的个人困扰及其转折点时刻。诠释互动论对人的体验及其回溯推理进行批判性的、基于存在的(existential)、诠释性的研究。这一视角遵循赖特·米尔斯的指引,他在《社会学的想象力》(*The Sociological Imagination*, Mills, 1959)中指出学者应该考察发生于直接的体验世界中的个体私人困扰是如何与公共议题相联系的,以及如何与对这些困扰的公众反响相联系的。米尔斯的社会学想象力是传记性的、互动性的、历史性的。诠释互动论进行的正是米尔斯的事业。

诠释互动论致力于让读者理解普遍人充满困扰的生命体验【Lived Experience】。互动论者解释这些世界及其影响、意义和再现。这种路数的研究方法包括表演文本;自传式民族志【Autoethnography】;诗歌;虚构文学;开放式的、创造性的访谈;文档分析;符号学【Semiotics】;生活史【Life History】;生活故事;个人体验与自我故事构建;参与观察【Participant Observation】;浓描【Thick Description】。诠释互动论这一术语表明了它是联结传统的符号互动论【Symbolic Interactionist】思想和诠释性探究的批判方式的一种尝试,这些诠释性探究的批判方式包括反思性的参与观察、文学民族志、诗学【Poetics】、生活故事和证词【Testimonios】。

诠释互动论不是所有人都喜欢。它的研究哲学基础与社会科学中大多数传统的科学研究惯例背道而驰。只有那些倾向于定性诠释方法的人更可能使用其方法与策略。这些策略赞成尽可能少地使用理论,赞成用精心编撰的文本来展示事实,而不是讲大道理(show, rather than tell)。

诠释互动论关注那些彻底改变和形塑了人们赋予其自身和生活的意义的生活经历。这种基于存在的要旨引导研究者关注"顿悟(epiphany)"。在顿悟中,个人性格得

以彰显。通过详尽研究这些经历,研究者能够阐明个人生活中发生的关键时刻。这些时刻常常被本人和他人加以诠释,正如转折点经历一样。有过这一体验之后,个人就再也不会有相同的体验了。诠释互动论研究这些经历,包括其回溯推理、影响、表演,以及与更宽广的历史、文化、政治和政治动力的关联。

——Norman K. Denzin

(高勇译校)

参考文献

Denzin, N. K. (1989a). *Interpretive biography*. Newbury Park, CA: Sage.

Denzin, N. K. (1989b). *Interpretive interactionism*. Newbury Park, CA: Sage.

Denzin, N. K. (1997). *Interpretive ethnography*. Thousand Oaks, CA: Sage.

Mills, C. W. (1959). *The sociological imagination*. New York: Oxford University Press.

诠释主义(Interpretivism)

诠释主义用来指涉一些有着共同的本体论【Ontology】和认识论【Epistemology】假定的社会科学研究路数。其核心原则是,由于自然科学和社会科学的主题存在根本差异,社会科学中不能使用自然科学的方法。研究社会现象需要理解人们栖居其中的社会世界,人们用他们生产和再生产出来的意义已经对社会世界进行了诠释,这种诠释是他们日常活动中必不可少的一部分。研究自然现象需要科学家使用科学概念和理论来诠释自然,需要科学家决定什么与研究问题相关,而社会科学家研究的现象是已经被人们诠释过的现象。

起源

诠释主义根植于德国的诠释学【Hermeneutics】和现象学【Phenomenology】的学术传统中,特别是马克斯·韦伯(Max Weber,1864—1920)和艾尔弗雷德·舒茨(Alfred Schütz,1899—1959)的著作。如与他们之前的狄尔泰一样,他们致力于建立一门以主观为研究对象的客观科学,目标是能够得到有关构成社会世界的意义的可验证的知识。他们的关注点是有意义的社会行动的本质,它在理解社会生活模式中的作用,以及如何评估这种意义。

韦伯和舒茨并未致力于确立某一社会行动者赋予某一具体社会行动的实际意义,他们认为有必要在更高的普遍性层次上进行工作。通过构建在典型情境下、在典型行动进程中、由典型社会行动者使用的典型意义的模型,就可以理解和解释社会的规律性。此类模型构成了要检验的尝试性假设。只有那些在性质上是理性的社会行动,那些被有意识地选择作为实现目标的方式的社会行动,才被认为是可以理解的。

根据韦伯(Weber,1964, p.96)的论述,主观意义可以分为三种类型:可以指涉社会行动者使用的实际的、有意图的意义;许多社会行动者使用的平均意义或近似意义;一个假想的社会行动者使用的典型意义(typical meanings)。

在借鉴狄尔泰的基础上,韦伯区分了三种理解类型:两种大类——理性的(rational)理解得到的是对社会行动在其意义语境中清晰而理智的了解,移情的(empathetic)或领悟的(appreciative)理解涉及对行动的情感内容的把握——和理性理解的两种小类,直接理解和动机理解。对人

的表达或行为的直接理解,类似于了解一个句子、念头或数学公式的意义。它是发生于日常情境中当下的、明晰的、就事论事的理解,不需要了解更为宽广的背景。另一方面,动机理解关心的是为了达到特定目标而选择何种手段。

韦伯主要关心的是理性行动的动机形式。他认为缺乏这种理性特点的人类行动是难以理解的。定量资料产生的统计模式,如教育获得和职业地位之间的关系,本身是不可理解的。不仅需要澄清连接关系中两个组成部分的相关行动,而且要说明附着于这些行动上的意义。

韦伯承认动机既可能是理性的,也可能是非理性的,它们既可能使得行动特征是为达成某一目标而采取的手段,也可能与情感品位有关。他认为只有在理性动机的情况下,进行社会科学的解释才是有可能的。

韦伯认为意义诠释是需要进行检验的可能假设,他推荐使用比较研究【Comparative Research】来对这些假设进行调查。但是,如果不可能进行比较研究,他认为研究者将不得不求助于“假想实验”,即设想动机链条上的特定要素,再构想出可能出现的行动过程。

韦伯主要关切的不是行动者自己的主观意义,而是对一个构建出的假想行动者而言的情境的意义。他特别感兴趣的也不是社会行动者赋予自身行动的具体意义,而是以近似和抽象方式得到的意义。他不只关心意义对人类行动的影响,而且关心有意义的社会关系对人类行动的影响。他的终极关切是大规模的规律性(large-scale uniformities)。因为他不希望局限于当代情境或微观情境,所以他必须处理观察者的诠释。但是,他认为这些诠释只是需要检验的假设。

韦伯企图将统计规律性和理解【Verstehen】(诠释性理解)勾连起来,这种企图导致对其诠释主义出现了多种诠释。那些重视其著作的实证主义者往往会把理解视为一种得到假设的可能方式,而诠释主义者往往会忽略他对统计规律和因果解释的关切。

因为韦伯从未脱离其实质性著作论述过方法议题,所以他的著作带有心照不宣的假设,他的某些概念也并未充分展开。韦伯著作中的这些方面已经被舒茨(Schütz,1976)系统性地接续,韦伯和胡塞尔为舒茨的著作奠定了基础。舒茨的主要目标是将韦伯的社会学置于坚实的基础之上。在完成这一任务的过程中,他提供了理想类型的方法。舒茨认为,韦伯主张有意义的行动这一概念是社会现象的基础而且是不可化约的组成部分,但韦伯没有能够区分社会行动者在行动发生之时携有的意义,社会行动者给予一个已经完成行动或某些未来行动赋予的意义,社会学家对行动赋予的意义。在第一种情况中,行动发生之时携有的意义和行动发生时的语境,通常都被视为理所当然的。在第二种情况中,赋予的意义将依据社会行动者的目标而给出。在第三种情况中,意义的语境是观察者的,而不是社会行动者的。韦伯似乎认为,后者是获得社会行动者赋予意义的充分基础,行动者之间以及行动者和观察者之间对已经完成的行动或未来的行动的意义不存在争论。

舒茨提供了连接社会行动者赋予的意义和社会科学家为了产生完善理论而赋予的意义之间的桥梁基础。他认为,社会理论中使用的概念和意义必须从社会行动者使用的概念和意义中得出。在舒茨的诠释主义版本中,日常现实是至高无上的;他认为对社会生活进行任何解释的基础都是社会行动者的观点,而非社会研究者的观点。韦伯和舒茨的著作为诠释主义的主要分支提供了基础。

批评

对诠释主义的批评既有来自这种方法

内部的,也有来自这种方法外部的。例如,吉登斯(Giddens, 1984)提出,认为社会行动者一直监控自己的行为,并因此能意识到自己的意图和行动原因,这种观点是误导性的。相反,通常只有在行动者对其行动进行反思性探寻时,在其他人质询其行动时,在行动被干扰或打断时,这种反思才会发生。多数时间里,行动的进行中是没有反思性监控的。

有些诠释主义者认为社会科学家必须只汇报社会行动者对其行动的解释而不对其加以阐释和理论化,雷克斯(Rex, 1974)及其他人则对此回应说,社会科学家对社会行动者的行动给出的解释可以不同于行动者自己给出的解释。巴斯卡尔(Bhaskar, 1979)将此称为语言谬误(linguistic fallacy),其根源在于没有能够认识到现实存在要比社会行动者用语言表达的更为丰富。巴斯卡尔(Bhaskar, 1979)和乌思怀特(Outhwaite, 1987)认为,诠释主义同时也犯了认识谬误(epistemic fallacy),因为诠释主义假定我们只能通过理解诠释过程才能进入社会世界中,假定诠释过程是全部的存在。其他批评者,如吉登斯和雷克斯认为诠释主义者没有能够认识到制度结构的作用,特别是利益分化和权力关系的作用。

——Norman Blaikie

(高勇译校)

参考文献

Bhaskar, R. (1979). *The possibilities of naturalism: A philosophical critique of the contemporary human sciences*. Brighton, UK: Harvester.

Blaikie, N. (1993). *Approaches to social enquiry*. Cambridge, UK: Polity.

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. Cambridge, UK: Polity.

Outhwaite, W. (1987). *New philosophies of social science: Realism, hermeneutics and critical theory*. London: Macmillan.

Rex, J. (1974). *Sociology and the demystification of the modern world*. London: Routledge & Kegan Paul.

Schütz, A. (1976). *The phenomenology of the social world* (G. Walsh & F. Lehnert, Trans.). London: Heinemann.

Weber, M. (1964). *The theory of social and economic organization* (A. M. Henderson & T. Parsons, Trans.). New York: Free Press.

四分位距(Interquartile Range)

对于已经从小到大排好序的定序数据来说,四分位距就是数据中的第 25 百分位数【Percentile】(P_{25})和第 75 百分位数(P_{75})之间的数值范围。由于这个范围中包括了全部数值中的一半,它是对分布【Distribution】的中间数值的描述。这个范围被称为四分位距,因为第 25 百分位数也被称为第一个四分位数【Quartile】,第 75 百分位数也被称为第三个四分位数。

试考虑如下有序数据集{3, 5, 7, 10, 11, 12, 14, 24}。这个数据集中有 8 个数值。去掉最前面的两个(即最前面的 25%)和最后面的两个(即最后面的 25%),剩下中间的 50%。其范围为 7~12,这就是这 8 个观察值的四分位距;但是,因为 P_{25}必须是大于 25%且小于 75%的数值,所以通常将 P_{25}定义为位于边界值(即 5 和 7)之间的中间值。这里就是 6。与此类似,第 75 百分位

数也是边界值 12 和 14 的中间值 13。因此,这个样本数据集的四分位距为 13-6,即 7。

用四分位距来描述数据,这丰富了用诸如中位数【Median】(即第二个四分位数)和算术均值【Mean】等点值进行的集中趋势描述。但是,四分位距的最大用处是生成盒形图【Boxplot】,以及用于定义游离值【Outlier】,即那些用定量方式被定义为极其异常的数据值。一种被普遍接受的游离值定义是,位于 P_{75} 之上或者 P_{25} 之下超过1.5倍四分位距的数据值(Tukey, 1977, p. 44ff)。在箱形图中,这些值用特殊的记号标注,通常是用星号。在前面的数据集中,24 就是一个游离值,因为它比 P_{75} 再加上1.5×(13-6)= 10.5还要大。

——Kenneth O. McGraw

(高勇译校)

参考文献

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

评判一致性(Interrater Agreement)

研究者常常会试图评估一组评审者(或评判人、观察者、专家、诊断工具)给出的评判或决策的共识度或一致性。评判的性质可以是定类的(如共和党/民主党/中间派,或是/否);定序的(如低、中、高);定距的;或定比的(如对象 A 的重量是对象 B 的两倍)。无论评判的类型是什么样的,评判一致性的共同目标是去评估评判者对对象的某一属性或某些属性的精确值在多大程度上是一致的;换言之,他们的评判在多大程度上是可以互换的。在实践和研究中,评判一致性往往会与评判信度【Interrater Reliablity】相混淆(Kozlowski & Hattrup, 1992; Tinsley & Weiss, 1975)。这两个术语代表不同的概念,需要不同的测量指标。例如,三个评审者 A,B,C 从四个维度来评判一本书稿:写作的清晰性;文献综述的全面性;方法的完善性;对此领域的贡献。给出的评判结果是六分类的:从 1(不可接受)到 6(优异)。评审者 A,B,C 在这四个维度上的打分结果分别是(1, 2, 3, 4)、(2, 3, 4, 5)、(3, 4, 5, 6)。数据清楚地表明,这些评审者的打分彼此完全不一样(即评判一致性为零),但是他们的打分在所有维度上是完全吻合的(即完美的评判信度)。

有许多种评判一致性的指标,它们是为不同的情况设计出来的,如不同的评判类型(定类、定序等)和不同的评判者数量或要评判的对象属性数量。这些指标的例子包括百分比一致性;评判标准差【Standard Deviation】;均值标准误【Standard Error】;斯科特(Scott, 1955)提出的 π;科恩(Cohen, 1960)提出的 kappa(κ)和加权 kappa(κ_w);劳里斯和路(Lawlis & Lu, 1972)提出的 χ^2 检验;劳希(Lawshe, 1975)提出的内容效度比(CVR)和内容效度指数(CVI);廷斯利和韦斯(Tinsley & Weiss, 1975)提出的 T 指数;詹姆斯、达马利和沃尔夫(James, Demaree and Wolf, 1984)提出的 $r_{WG(J)}$;以及林德尔、勃兰特和惠特尼(Lindell, Brandt & Whitney, 1999)提出的 $r^*_{WG(J)}$。在下面的部分,我们将介绍斯科特的 π,科恩的 κ 和 κ_w,廷斯利和韦斯的 T 指数,劳希的 CVR 和 CVI,詹姆斯等人的 $r_{WG(J)}$ 和林德尔等人的 $r^*_{WG(J)}$。这些指标都被广泛地应用,其相应的虚无假设【Null Hypothesis】也已经非常成熟。斯科特的 π 适用于两个独立评判者对

一组评判对象的某一属性进行定类决策。其定义为 $\pi=\frac{p_o-p_e}{1-p_e}$，实际取值范围为 0～1（虽然也可能得到负值），其中的 p_o 和 p_e 分别代表观察到的两个评判者一致的百分比和由于概率而期望得到的一致百分比。p_e 是由评判类别数量和两个评判者使用的类别频率决定的，定义为全部类别的比例平方的加总：

$$p_e = \sum_{i=1}^{k} p_i^2$$

其中的 k 为类别总数，p_i 为评判结果落入第 i 个类别的比例。

科恩的 κ 是被最广泛使用的指数，它评估两个或多个独立评判者对某一属性进行定类决策时的一致性。类似于斯科特的 π，科恩的 κ 的定义为 $\kappa=\frac{p_o-p_e}{1-p_e}$，取值范围通常为 0～1。但是，科恩的 κ 中的 p_e 以不同的方式加以操作化，它是所有类别的列联表【Contingency Table】的边缘【Marginals】的联合概率的加总。不同于科恩的 κ，科恩的 κ_w 将不一致的情况也考虑进来。在诸如人事选择等某些实践情况中，“肯定能成功”及“可能会成功”的不一致，和“肯定能成功”及“肯定要失败”的不一致比起来，显得更无足轻重。在评判者数目增加时，κ 和 κ_w 都会过分地估计出现一致的概率。

廷斯利和韦斯的 T 指数，$T=\frac{N_1-Np_e}{N-Np_e}$，是基于科恩的 κ 和劳里斯及路 χ^2 对于评判随机性【Randomness】的卡方检验而发展出来的。其中的 N_1 是评判一致的属性数目、N 是属性总数目，p_e 是由于随机因素而出现的期望一致概率。上述公式除以 N，可以重新写作 $T=\frac{p_o-p_e}{1-p_e}$，其中的 p_o 是评判者一致的属性比例。p_e 取决于可接受的评分范围（r），用来评判各个属性的评分类别数目（A），评判者数目（K）。如果研究者要求完全一致（即评分范围为 0），$p_e=\left(\frac{1}{A}\right)^{K-1}$。如果 $r>0$，p_e 的计算公式会很复杂；结果，从廷斯利和韦斯那里会得到许多个 p_e 的值（1997，注 7）。

劳希的 CVR 最初是用来评估专家对于测试题项重要性的一致性的，其定义为 $\text{CVR}=\frac{p_A-0.5}{0.5}=2p_A-1$，取值范围从 -1～1，其中的 p_A 为达成一致的评判者比例。如果评判者对多个题项进行判断，就要用到整体评判一致性指数（简称为 CVI）。CVI 的定义为 $\text{CVI}=2\bar{p}_A-1$，其中的 $\bar{p}_A$ 是各个题项的 p_A 值的平均数。

詹姆斯等人提出了 $r_{WG(J)}$，它比较的是一组评分的观察方差和源自随机打分的期望方差之间的差别。如果评判的只是一个属性，则：

$$r_{WG(1)} = 1 - \frac{S_x^2}{S_E^2}$$

其中的 S_x^2 是对某一属性 X 的评分的观察方差，S_E^2 是随机打分的方差。虽然有许多种方法来决定 S_E^2，但研究者通常假定随机打分的分布是服从离散均匀分布的（Cohen, Doveh, & Eick, 2001）。在这种情况下，可以用 $S_{EU}^2=\frac{A^2-1}{12}$ 来估计 S_E^2，其中的 A^2 是对属性 X 进行评判的类别数目。如果对 J 个属性进行评判，则有：

$$r_{WG(J)} = \frac{J\left(1-\frac{\bar{S}_x^2}{S_E^2}\right)}{J\left(1-\frac{\bar{S}_x^2}{S_E^2}\right)+\frac{\bar{S}_x^2}{S_E^2}}$$

其中的 $\bar{S}_x^2$ 是 J 个属性的平均方差。类似

地，如果假定随机打分的分布服从均匀分布，则 S_E^2 可以用来 S_{EU}^2 估计。类似于前面的指标，$r_{WG(J)}$ 的取值上限为 1。如果 $r_{WG(J)}$ 取值为负，詹姆斯等人建议将其设定为零。

詹姆斯等人提出的 $r_{WG(J)}^*$ 与詹姆斯等人的 $r_{WG(J)}$ 在两个方面有所不同。首先，他们认为大于随机期望的评判一致和小于随机期望的评判一致是同样成立的（Lindell & Brandt, 1999, p. 642）。科恩等人（Cohen, Doveh, & Eick, 2001）发现，如果把 $r_{WG(J)}$ 的负值设定为零，就会出现正向的偏差。其次，詹姆斯等人的 $r_{WG(J)}$ 源自于斯皮尔曼-布朗【Speraman-Brown】相关公式，虽然 $r_{WG(J)}$ 是一个一致性指标而不是一个信度指标。林德尔等人质疑对一致性指标使用斯皮尔曼-布朗修正的合理性，他们建议在均匀分布假定下可以使用如下公式：

$$r_{WG(J)}^* = 1 - \frac{\bar{S}_x^2}{S_E^2}$$

或者

$$r_{WG(J)}^* = 1 - \frac{\bar{S}_x^2}{S_{EU}^2}$$

总体而言，$r_{WG(J)}$ 和 $r_{WG(J)}^*$ 在许多情形下都比劳希的 CVR 和 CVI、廷斯利和韦斯的 T 指数更为灵活。

——Peter Y. Chen
Autumn D. Krauss
（高勇译校）

参考文献

Cohen, J. (1960/1968). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37-46.

Cohen, A., Doveh, E., & Eick, U. (2001). Statistical properties of the $r_{WG(J)}$ index of agreement. *Psychological Methods*, *6*, 297-310.

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, *69*, 85-98.

Kozlowski, S. W., & Hattrup, K. (1992). A disagreement about within-group agreement: Disentangling issues of consistency versus consensus. *Journal of Applied Psychology*, *77*, 161-167.

Lawlis, G. F., & Lu, E. (1972). Judgment of counseling process: Reliability, agreement, and error. *Psychological Bulletin*, *78*, 17-20.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*, 563-575.

Lindell, M. K., & Brandt, C. J. (1999). Assessing interrater agreement on the job relevance of a test: A comparison of the CVI, T, $r_{WG(J)}$, and $r_{WG(J)}^*$ indexes. *Journal of Applied Psychology*, *84*, 640-647.

Lindell, M. K., Brandt, C. J., & Whitney, D. J. (1999). A revised index of interrater agreement for multi-item ratings of a single target. *Applied Psychological Measurement*, *23*, 127-135.

Scott, W. A. (1955). Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly*, *19*, 321-325.

Tinsley, H. E., & Weiss, D. J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology*, *22*, 358-376.

评判者信度（Interrater Reliability）

行为科学家常常需要评估评判者对某些特征的判断是否一致，如判断老年病房中

病人的攻击性行为程度、文稿的质量、花样滑冰运动员的艺术表现力、治疗师的诊断能力等。评判者之间的判断不一致,这会威胁从这些判断中得出的任何推断的效度。评判者信度测量评估的就是两个或多个评判者的判断的相对一致性(relative consistency)。

两个评判者的判断可以看成一次测试中两种备择试卷上的分值。因此,根据经典的信度【Reliability】理论,对两个评判者的信度就可以用简单相关【Simple Correlation】来进行估计,如皮尔逊相关系数【Pearson's Correlation Coefficient】或者斯皮尔曼相关系数【Spearman Correlation Coefficient】。但是,如果评判者超过两个,这些方法就没有用了,例如,三位军官对每一架 F-18 战斗机的着陆表现进行了评分,此时一个海军指挥官需要估计他们的评分是否一致。如果每架战斗机的评分是由不同组官员给出的,那么上述方法也不能充分评估评判者信度。在上述情况下评估评判者信度,最佳的测量是组内相关【Intraclass Correlation,ICC】,这是一个单侧面的(one-facet)概化【Generalizability】研究的特例。

在概念和理论上,组内相关系数的取值范围为 0~1(虽然实际上它并不总是如此),它是一个比值,分子是与被评判者有关的方差,分母是与被评判者有关的方差与误差方差的和。不同的情形中用到不同的 ICC 公式。我们将主要关注三种在实践和研究中都会经常遇到的情形。对于更为复杂的情形,如对四个评判者在三个不同时点上的判断进行评判者信度评估,就应当采用概化理论来处理了。

什鲁特和弗莱斯(Shrout & Fleiss,1979)在一篇开创性的文章中说明了这三种情形:(a)随机样本中的每一个被评者被不同的一组评审者进行独立评判,评审者是从总体中随机抽选的;(b)随机样本中的每一个被评者被相同的一组评审者样本进行独立评判,评审者是从总体中随机抽选的;(c)随机样本中的每一个被评者被不同的一组评审者进行独立评判,评审者只包括特定的人(也就是说,我们并不关心结果是否可以推广到评审者总体中)。如下所示,每种情形下计算相应的 ICC 都需要不同的统计分析。具体而言,情形 1、情形 2、情形 3 中合适的模型分别为一维随机效应模型、二维随机效应模型、二维混合模型。注意,所有这些 ICC 指标都是有偏但一致的信度估计。

表 1　用于情形 1、情形 2、情形 3 下 ICC 计算的示例数据

被评者	评审者 1	评审者 2	评审者 3
A	2	4	7
B	2	4	7
C	3	5	8
D	3	5	8
E	4	6	9
F	4	6	9
G	5	7	10
H	5	7	10
I	6	8	11
J	6	8	11

表 2　情形 1 中的方差分析结果

方差来源	*df*	*MS*
被评者之间	9	6.67
被评者内部	20	6.33

在情形 1 中,假定有 10 个被评者,每个被评者被一组不同的评审者评判。例如,给被评者 C 评分的三个评审者与给被评者 E

评分的三个评审者是不同的人。基于表 1 中的数据和表 2 中的方差分析【Analysis of Variance】结果,可以按如下公式计算 ICC:

$$ICC(1,1)=\frac{MS_b-MS_w}{MS_b+(k-1)MS_w}$$

其中的 MS_b 是被评者之间的均方,MS_w 是被评者内部的均方,k 是评审每个被评者的评审者人数。如果评审每个被评者的评审者人数不同,k 可以用评审每个被评者的评审者人数均值来估计。ICC(1,1)是单个评审者评分的期望信度,结果为 0.02,表明被评者仅仅能够解释分数中 2%的变异。如果我们对于 k 个评审者平均分的评判者信度感兴趣,可以用下式进行估计:

$$ICC(1,k)=\frac{MS_b-MS_w}{MS_b}$$

结果为 0.05。

在情形 2 中,10 个被评者被相同的 3 位评审者评判。这些评审者是从评审者总体中抽取的随机样本。根据表 1 和表 3,我们可以用下式计算 ICC:

$$ICC(2,1)=\frac{MS_b-MS_e}{MS_b+(k-1)MS_e+\frac{k(MS_j-MS_e)}{n}}$$

其中的 MS_b 是被评者之间的均方,MS_j 是评审者之间的均方,MS_e 是均方误差。k 是评审者人数,n 是被评者人数。基于样本数据,ICC(2,1)为 0.35。如果我们对于 k 个评审者平均分的评判者信度感兴趣,可以用下式进行估计:

$$ICC(2,k)=\frac{MS_b-MS_e}{MS_b+\frac{MS_j-MS_e}{n}}$$

结果为 0.51。

表 3 情形 2 和情形 3 中的方差分析结果

方差来源	*df*	*MS*
被评者	9	6.67
评审者	2	63.33
残差	18	5.403E-15

在最后一种情形中,同样 3 位评审者对 10 个被评者进行评判。情形 2 和情形 3 之间的差别在于,情形 2 中我们对将结论推广到评审者总体中感兴趣,而情形 3 中我们只对实际的评审者感兴趣。类似地,我们可以利用表 3 中的信息,按下式计算 ICC:

$$ICC(3,1)=\frac{MS_b-MS_e}{MS_b+(k-1)MS_e}$$

结果 ICC(3,1)接近 1。如果我们对 k 个评审者平均分的评判者信度感兴趣,只要评审者和被评者之间不存在交互影响,就可以用下式进行估计:

$$ICC(3,k)=\frac{MS_b-MS_e}{MS_b}$$

——Peter Y. Chen
Autumn D. Krauss
(高勇译校)

参考文献

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer.* Newbury Park, CA: Sage.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420-428.

中断时间序列设计(Interrupted Time-Series Design)

中断时间序列设计可以估计出一个离散型干预的因果【Causal】效应。在其最简单的形式中,研究始于对一个因变量【Dependent Variable】在长时段中进行一系列重复测量。干预【Intervention】将此时间序列【Time-Series】划分为干预前和干预后两个时期,数据分析要比较在这两个时期中因变量的均值。如果均值间没有差异的虚无假设【Null Hypothesis】被拒斥,这就被认为证明了干预确实对这一序列存在影响。

图1就是上述设计逻辑的一个极其清晰的例子。数据来自约翰·麦克斯威尼在1978年的研究,是俄亥俄州辛辛那提市180个月的本地查号台呼叫数据。在1974年3月,辛辛那提电话公司开始向这些呼叫收取少量费用,在此之前这些业务是免费的。呼叫量直线下降,说明费用成功地降低了对这项业务的需求。



图1　辛辛那提市1962年1月—1976年12月本地平均每天的查号呼叫(单位:千)

来源:McSweeny(1978)。

在这个例子中,干预的因果效应看起来很明显。呼叫的数量迅速大幅下滑,也没有其他解释能够说明为什么这种下降会发生。更常见的情况是,序列中的变化看起来并不明显,也更容易想出另外一些备择解释。因此,运用时间序列设计必须要处理统计分析问题,研究者必须考虑其他备择解释来证明表面的因果效果确实存在。

这种设计的流行很大程度上源自唐纳德·坎贝尔的著作,其著作分别是与朱利安·斯坦利(Campbell & Stanley, 1963)和托马斯·库克(Cook & Campbell, 1979)合作完成的。坎贝尔及其合作者列出了一系列损害因果推论的内在效度的因素清单,在这些情况中真正能解释结果的是研究之外的那些变量。坎贝尔及其合作者用他们的清单来评估通常的非实验设计,最后得出结论认为中断时间序列设计的内在效度比大多数其他方法都高。

尽管时间序列设计有其相对优势,但它仍然容易受到几个重要的效度问题的影响。其中一个就是历史,或者在干预时点前后发生的其他事件可能导致了观察到的效应。为了证明历史的影响并不能说明问题,研究者通常会分析没有接受干预的控制序列(control series)。这些序列中没有发生变化,会使得我们更有信心认为干预是真正原因所在。例如,在辛辛那提,电话公司继续提供免费的长途查号服务,在本地查号服务收费之后,长途查号服务的序列并没有出现下降。

对中断时间序列设计的统计分析通常依赖于Box-Jenkins建模法【Box-Jenkins Modeling】。Box-Jenkins方法虽然并不是绝对必需,但是它有助于解决两个主要的分析问题:非平稳性和序列相关【Serial Correlation】。非平稳性发生于序列缺少稳定的均值之时(即它具有趋势时),序列相关发生于后期观察可以用前期观察来预测

之时。如果不解决这些问题,它们就会使得对干预前和干预后的均值差异比较的检验变得无效。

——David McDowall

(高勇译校)

参考文献

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton-Mifflin.

McDowall, D., McCleary, R., Meidinger, E. E., & Hay, R. A., Jr. (1980). *Interrupted time series analysis*. Beverly Hills, CA: Sage.

McSweeny, A. J. (1978). Effects of response cost on the behavior of a million persons: Charging for directory assistance in Cincinnati. *Journal of Applied Behavior Analysis*, *11*, 47-51.

定距(Interval)

定距是一种测量层次,在此测量层次下变量的数值是等距的,而且有内在意义。典型的例子就是用美元单位测量的收入。收入值为 90 000 美元,这很好理解。此外,收入为 90 000 美元的人和收入为 90 001 美元的人之间的差距(即一个美元),等于收入为 56 545 美元的人和收入为 56 546 美元的人之间的差距(也是一个美元)。定距的测量层次具有较高的精确度,可以现成地使用多元统计学。定距测量有时也被称为定量【Quantitative】测量,或者计量【Metric】测量。定距数据可以转换成为更低的测量层次,如定序【Ordinal】或定类【Nominal】的,但是会损失信息。

——Michael S. Lewis-Beck

(高勇译校)

干预变量(Intervening Variable)

参见中介变量【Mediating Variable】。

干预分析(Intervention Analysis)

在时间序列中的干预测度,代表着在一个时间段中的系列数据出现了断裂。此类概念化的目的常常受评估政策的影响,或者评估历史过程中的其他突变(discrete introductions)的影响。在许多情况下,它们是作为中断时间序列设计【Interrupted Time-Series Design】中的实验处理来引入的(McDowall, McCleary, Meidinger, & Hay, 1980)。对政府干预或其他影响时间序列的因素的估计可能会受到一些污染因素(contaminating factors)的影响而是偏倚【Bias】,如数据中存在趋势【Trend】(非稳定

性）、自回归【Autoregression】（即使不存在趋势时也可能存在，如在循环过程中）、移动平均【Moving Average】的部分（长时间存在然后逐渐变弱的随机冲击）等，因此分析模型通常设定为干预存在与否（0或1），再加上一个已经去除了与时间相关的任何影响的噪声模型——一个白噪声【White Noise】过程。这个过程通常表示为如下公式：

$$y_t = \omega I_t + N_t$$

其中 y_t 表示时间序列中的一个观察，它由干预效应 ωI_t 和一个白噪声过程（white noise process）N_t 组成。干预效应可以通过对 ω 显著性的 t 检验【t-Test】来进行评估。这样做的效果是去除了干预效应之外的时间序列因素，因此可以对政策影响或其他干预有关的假设进行检验。

虽然研究采用了坎贝尔和斯坦利（Campbell & Stanley, 1966）建议的准实验【Quasi-Experiment】设计形式（参见中断时间序列设计【Interrupted Time-Series Design】），但是干预效应的估计可以采用许多设定方式。如果可以获知干预形式的确切信息，如可允许的污染量每年减少（Box & Tiao, 1975），那么干预可以进行相应设定。但是，大多数政策干预是很难设定的，需要用虚拟变量【Dummy Variable】来表示这项政策或项目存在与否。对于这些更一般的效应来说，存在三种基本的设定模型来解释干预（图1）。

通过在估计式中加入 y 的滞后值 y_{t-1}，然后用 t 检验来评估参数 δ，就可以检验上述模型：

$$y_t = \delta y_{t-1} + \omega I_t + N_t$$

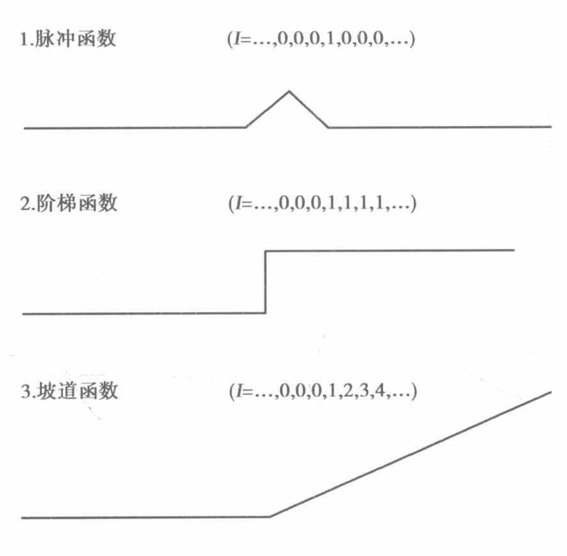


图1　干预的基本结构

如果分析者不是很清楚地了解干预形式，就应该对不同的设定进行检验。具体而言，可以用以下方式来尝试设定干预形式：

（1）如果设定的脉冲函数（pulse function）（突然的即时效应）的参数显著，而参数 δ 并不显著，干预效应的形式应为图1中所示的脉冲函数。

（2）如果设定的脉冲函数（突然的即时效应）的参数 ω 显著，而参数 δ 也是显著的，这种结果表明阶梯函数（突然的永久效应）是更好的设定。例如，如果滞后内生变量的系数为1，模型等同于一个阶梯函数（step function）。

（3）如果设定的阶梯函数的参数 ω 显著，但是参数 δ 并不显著，阶梯函数的设定可能就是正确的（图1）。

（4）如果设定为阶梯函数的干预参数和滞后时间序列的参数 y_{t-1} 都是显著的，就应当考察其他的模型组合。

（5）如果脉冲函数的参数不显著，而 y_{t-1} 的参数显著，就应该考察坡道函数（ramp function）（图1）。

如果在前述模型设定中发现了模棱两可的情况，模型就可能需要更复杂的其他设

定,有时是基本模型的组合(图 2)。

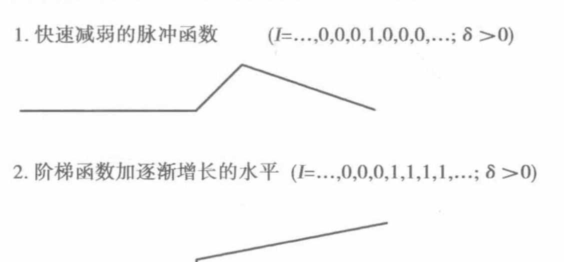


图 2 干预效应的常见复杂模型

这些干预模型代表了最常见的设定。更为复杂的模型,包括多重干预,在代表政策效应和项目效应中也是有用的。

学者们有时会提出反对意见,认为遵循博克斯和詹金斯(Box and Jenkins,1976)对白噪声过程的设定方法,把与时间相关的过程去除掉,这种做法是罔顾理论的。如果研究者相信可以设定出变量代表其他更可信的假设,就可以在模型中包括这些变量作为转移函数。最为理论化的过程就是所谓的误差修正模型【Error Correction Model】,它能够从方程中去除很多混杂其中的统计假象(statistical artifacts),如共线性【Collinearity】。

——Robert B. Albritton

(高勇译校)

* 也可参见中断时间序列设计【Interrupted Time-Series Design】。

参考文献

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.

Box, G. E. P., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, 70, 70-92.

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Skokie, IL: Rand McNally.

McDowall, D., McCleary, R., Meidinger, E. E., & Hay, R. A., Jr. (1980). *Interrupted time series analysis*. Beverly Hills, CA: Sage.

访谈提纲(Interview Guide)

访谈提纲,或访谈备忘录(aide memoire),就是在半结构化访谈【Semistructured Interview】中要涵盖的论题、主题或领域的清单。研究者通常会在访谈前制订访谈提纲,制订的方式应当在涵盖的论题与领域、对每个被访者的提问方式、问题的顺序方面保持灵活性和流畅性。访谈提纲通常会和指导研究的研究问题【Research Question】有关联,涉及的领域应当能够得到有助于解决这些问题的资料。

访谈提纲是用来帮助访谈者进行有效的半结构化访谈的工具。它可以采用多种形式,但是首要的考虑是它要帮助访谈者在访谈的社会互动过程中,在合适的时机,以合适于被访者的方式,提出与研究关注点相关的问题。一个访谈提纲应当帮助访谈者对访谈内容和访谈顺序进行现场决策。这意味着要决定被访者引入的一个话题是否值得跟进,什么时候结束一个话题再开始一个新话题,要不要以及什么时候重提访谈中已经提及过的事情,什么时候以及怎么去追问更多的细

节。这也通常意味着根据具体访谈的具体情境,现场决定如何措辞提问,而不是念稿子。

具有很高灵活性和很强针对性的访谈问题不是用访谈问题的标准化清单能够获得的,因此访谈提纲应当有别于访谈安排【Interview Schedule】,访谈安排中包括了对每个被访者都要问及的正式问题清单。与此不同,访谈提纲可以是便于洗牌和改变顺序的一些卡片,卡片上是论题的题目和关键词。它也不妨包括一些关键议题的提问方式的示例,但是它不会详细地列举每个问题甚至任何问题。它也可以包括一些研究者的直觉,或者包括那些一旦出现就值得追问的具体议题或事件。它也可以包括一份检查单,上面列举了在访谈结束时需要提问的领域或者被访者的特征,如果这些还没有被问到的话。

访谈提纲的质量取决于它对研究者和进行的研究的有效程度,而不是抽象的原则。因此至关重要的是,访谈提纲要在预研究【Pilot Study】中进行检验,然后进行相应修改。

——Jennifer Mason

(高勇译校)

参考文献

Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing*. Thousand Oaks, CA: Sage.

Mason, J. (2002). *Qualitative researching* (2nd ed.). London: Sage.

访谈安排(Interview Schedule)

访谈安排是访谈者在进行结构性访谈【Structured Interview】时使用的指导手册。它有两个部分:一系列问题被设计出来严格按照其措辞来进行提问,还有一系列指令来指导访员如何一步步地提出这些问题。

问题要按照提问的顺序安排。问题要设计得能够在提问中一字不差、逐字逐句地读出。问题不仅要告诉被访者【Respondent】问的是什么信息,而且要告诉被访者应该以何种方式或形式来回答这个问题。例如,"年龄"并不能构成访员提出的问题。在结构性访谈安排中,问题应当是访谈者所需要的所有字句:"请问上一个生日的时候,您多大年纪?"

与此类似,"您什么时候出生的?"这个问题并没有告诉被访者应该怎么回答。"越南战争之后""20 年前""1971 年",这些都是对这个问题的合理回答,但是却彼此不同。访谈安排中一个好的问题应当是这样的:"您出生在哪一年?"

访谈安排还需要说明访员如何一步步地提出这些问题。例如:

(1)提供备选措辞来贴合被访者的具体情况。例如,"您(或您的丈夫/妻子)在过去 12 个月里有没有买过股票?"括号中的"或您的丈夫/妻子"告诉访员,他们必须要让问题适合被访者的境况。这有助于访员措辞恰当地进行提问,并确保所有访员都能够以相同一致的方式来变换问题的措辞。

(2)提供跳答指令。例如,特定问题只向那些最近看过病的人提问。访谈安排必须分辨出那些曾去看过病和没有看过病的人来,而且要有合适的指令来让访员知道哪些问题该答,哪些问题该跳过。

(3)访谈安排也要有指令让访员知道在哪儿以及如何记录。

访谈安排可以是以纸本形式写在小册

子上。越来越多的访谈安排已经计算机化了,用于计算机辅助数据收集【Computer-Assisted Data Collection】;要问的问题在计算机屏幕上显示给访员阅读,答案记录方式要么是把被访者的话键入计算机,要么是输出对应于特定答案的一个数字。在两种形式中,访谈安排的作用在于,提供给访员一个规范协议【Protocol】,以对所有被访者一致的方式来提问和记录答案,促进访谈过程顺畅而高效地完成。

——Floyd J. Fowler, Jr.

(高勇译校)

参考文献

Converse, J., & Presser, S. (1986). *Survey questions*. Beverly Hills, CA: Sage.

Fowler, F. J., Jr. (1995). *Improving survey questions*. Thousand Oaks, CA: Sage.

Sudman, S., & Bradburn, N. (1982). *Asking questions*. San Francisco: Jossey-Bass.

访员效应(Interviewer Effects)

由于访员差异而使得样本【Sample】统计量的方差【Variance】增加,这被称为访员效应。可以用模型表明,增加的方差源自同一访员收集的回答之间的正相关【Correlation】系数 ρ_{int}。在这个简单的模型【Model】下,可以表明样本均值的方差将增加为 $1+(\bar{m}-1)\rho_{int}$ 倍,其中 $\bar{m}$ 是平均每个访员的工作量(Kish, 1962)。

研究表明,ρ_{int} 通常都很小,一般小于 0.02,几乎不会大于 0.10。即使如此,在每个访员工作量较大时,它仍然会对统计量的标准误差【Standard Error】造成很大影响。例如,ρ_{int} 为 0.02,平均每个访员的工作量为 150 份访谈时,方差将会增加为 $1+(150-1)\times 0.02=3.98$ 倍,题项的样本均值【Mean】的标准误差将会翻倍。这说明降低访员差异对调查数据影响的办法之一是增加访员数量,减少他们的平均工作量。

访员效应并不是调查操作中的常规测量部分。要测量访员效应,研究者要介入访员分配中,意味着样本单位必须要随机分配【Random Assignment】给他们。这在面访调查中成本惊人,但在电话调查操作中更加可行。即使如此,潜在的工作转换效应会使得估计【Estimation】过程很复杂。一种被称为多水平分析【Multilevel Analysis】的方法可以用来解释访员和受访者的协变量,这种方法已被用来为访员效应建立模型。现在也很容易得到实现这种方法的软件(如 MLwiN),这使得在实际工作环境中估计访员效应变得更为可行。

测量访员效应的好处之一在于可以更现实地评估均值估计和加总估计的精确性。另一个好处是,它提供了一种在不同调查题项之间评估和比较其数据质量的方法。理论上讲,这些信息有助于督导和提升访员的表现。但是,现在仍不清楚如何降低一个题项的 ρ_{int},即使已经知道它比较大。很多研究试图发现访员效应的原因所在。访员的人口和行为特征,如种族、性别、声音、风度,都被研究过了。没有证据表明这些特征能够解释访员效应。还有一些研究试图确定容易受到影响的题项类型。人们看似合理地认为敏感问题可能更易受影响,但这并没有得到确切证明。有证据显示,那些需要访员干预(如追问【Probing】)的题项有更大的访员效应,言下之意是对访员进行更多的技巧培训可能会有用。但是,另一些研究表

明,单单增加培训是没有用的,但是通过集中电话设备(centralized telephone facility)进行监控和反馈可能有用。直到最近,人们普遍认为标准化访谈是控制访员效应的最佳方法,它限制访员按统一问卷进行访谈。现在有一些研究表明,允许访员转述或澄清回答的更多对话式的风格,可能会增加回答的准确度,但并不会增加 ρ_{int} (Schober & Conrad, 1997)。

——S. Lynne Stokes

(高勇译校)

参考文献

Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A., & Sudman, S. (Eds.). (1991). *Measurement errors in surveys*. New York: Wiley.

Kish, L. (1962). Studies of interviewer variance for attitudinal variables. *Journal of the American Statistical Association*, *57*, 92-115.

Schober, M., & Conrad, F. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly*, *61*, 576-602.

访员培训(Interviewer Training)

社会科学研究项目中的访员几乎总是可以从培训中获益。他们需要学习的具体技能和技巧因他们要进行的是结构化访谈【Structured Interview】或半结构化访谈【Semi-Structured Interview】而有所不同。

如果访员要进行的是结构化访谈,他们需要学习的基本技能包括:

①完全逐字逐句地问问题。

②以非诱导方式对不完整的回答进行追问,即追问方式不要增加某一种回答相对于其他回答的可能性。

③原原本本地记录回答。

如果访员进行的是半结构化访谈,他们需要培训怎样说话或者怎样提问,以便得到他们需要获取的信息。对半结构化访谈的回答往往是用被访者自己的话来叙述的,而结构化访谈往往要求被访者从一组答案中选择一项。因此,半结构化访谈的访员必须有比结构化访谈的访员更高的技巧来进行非诱导性的追问。具体而言,这类访员需要培养和学会这些技巧,能够诊断出答案不完整的原因,以及能够提出一系列非诱导性的追问来得到所需的信息。

对于这两类访谈来说,下述两种技巧是共同的。首先,访员通常需要得到潜在被访者的合作。因此,他们需要学习一种能让被访者感兴趣并愿意接受访谈角色的方式介绍研究项目。其次,访员也必须培训被访者在访谈中应该怎么做。这尤其对结构化访谈是个挑战,因为其中访员和被访者的行为都是严格规定好的。对这类访谈,访员向被访者恰当地解释他们应当怎么做和为什么这样做最好是很重要的,这样可以让被访者接受和完成自己的角色。比较起来,半结构化访谈约束少一些,因此培训被访者接受和完成角色的挑战也就更小一些。

已经有数项研究表明,对结构化访谈的访员的培训对他们执行访谈的水平和收集到的数据的质量有重要影响。只接受了最低限度培训的访员表现得不好。尤其是研究显示有督导的访谈练习对访谈和数据质量有正面影响。关于培训对半结构化访谈获得资料的影响,我们还缺乏系统的研究。但是已有研究表明,访员最难掌握的技巧之

一就是对以叙述形式提供的回答进行追问【Probing】,这些技巧也是从培训中获益最多的。这些研究还表明,可能更难的技巧是提问的措辞要设计得体,而这也更可能从培训中获益。

——Floyd J. Fowler, Jr.

(高勇译校)

参考文献

Fowler, F. J., Jr., & Mangione, T. W. (1990). *Standardized survey interviewing*. Newbury Park, CA: Sage.

Weiss, R. (1995). *Learning from strangers*. New York: Free Press.

访谈(Interviewing)

在社会科学中,资料收集的一种主要方法就是对个人进行访谈。访谈可以很正式地以社会调查的形式实行、面对面或者通过电话进行。或者,一个研究总体中的对象,例如安第斯山脉中的土著社会成员,可以非正式地对其进行访谈。在精英层面,访谈基本上以对话形式进行,至少表面上看如此。访谈方法可能必须适应不同的社会语境,以便达到最重要的目标,即获得有效且有用的资料。

——Michael S. Lewis-Beck

(高勇译校)

定性研究中的访谈(Interviewing in Qualitative Research)

访谈【Interviewing】是社会科学中最常使用的研究方法之一,在定量抽样调查研究和定性研究【Qualitative Research】中都会用到。定性访谈是一种特殊的谈话,一个或多个访谈者对一个或多个被访者就某个或某些话题提出问题、仔细聆听、记录回答。有些访谈通过电话进行,有些当面进行;有些会录像,大多数则是录音然后再转写。以下对定性访谈的讨论基于最常见的情况:单个访谈者使用录音设备访问单个被访者。

访谈自19世纪以来就在社会科学中使用,那时学者和实际从业者开始试图回答关于人类境遇的大问题。在美国,1920年代和1930年代的芝加哥大学中,访谈作为社会科学中个案研究方法的一部分被使用。除此之外,还有实地研究、档案分析、统计方法(Platt, 2002)。第二次世界大战后,访谈开始在抽样调查研究中广泛使用;从此以后,可定量化的封闭式问题【Closed-Ended Question】调查访谈变得比开放式的定性访谈更为普遍。

但是,定性访谈仍然是人类学的重要方法,如今所有社会科学都在使用定性访谈。在社会科学研究中定性访谈的目的和普遍的定性研究一样,是理解访谈主题对于被访者的意义。如苛费尔(Kvale,1996)所言,定性访谈的目的是"得到被访者生活世界的描述,特别是对所描述现象的意义理解"(pp. 5-6)。为简明起见,我用我自己关于女性精神病患者(madwives)的书为例来说明这一点,这本书就是基于定性访谈的(Warren, 1987)。

在1950年代后期和1960年代早期对女性精神病患者的访谈中(Warren, 1987),

八位访谈者想了解在女性患者和精神病医生眼中的住院治疗体验。他们在住院期间和出院后隔周对这些女性进行访谈，有的访谈持续了数年之久。他们提及的问题之一是电击疗法或者电痉挛疗法(electroconvulsive therapy，ECT)。从医学模型的角度来看，ECT 是一种精神病治疗方法，意在减轻抑郁、精神分裂或者其他失调症状，会有包括短期失忆在内的副作用。但是，对于在 1950 年代后期和 1960 年代早期接受 ECT 的女性来说，这种疗法的意义不仅和医学模型的解释不同，而且也是不断变化的(因人而异，有时因时而异)。有些女性认为它是一种惩罚手段，其他人尽管相信它是一种医学疗法，但认为目的在于从记忆中消除困扰。

社会科学对人们进行访谈，是因为他们对被访者的生活世界心存疑问。精神病患者是怎样解释给他们的各种治疗方法的？与已婚男士相恋的单身女性如何看待其处境(Richardson，1985)？要使用定性访谈方法，这些研究问题【Research Question】就必须转换成研究者向被访者提出的实际问题。这些问题的产生要考虑两种背景：整体的研究设计【Research Design】和保护研究对象的政府规范和制度规范。

在社会科学中，定性访谈可以是包括实地研究【Field Research】或其他方法的研究设计中的一部分，也可以是研究者计划的唯一方法。无论整体设计如何，访谈者问的具体问题都要和感兴趣的主题以及研究设计者的研究问题相关。一般而言，定性研究者会想出 10～15 个问题来问被访者，他们也会用追问【Probes】来获取更多信息。为了节省时间，可以给每个被访者一张基本情况表(face sheet)来收集诸如性别、年龄、职业、民族等基本人口信息。这样，剩余的访谈时间就可用在研究者的主题问题和被访者的叙述上了，访谈时间长度从 15 分钟到几个小时都可以，这要取决于提问者以及被访者的健谈程度。

在当代社会科学研究中，定性访谈研究必须关注用来保护研究对象的政府规范和制度规范。虽然本卷书中其他地方讨论了以人为研究对象的研究规范，注意到访谈研究中的知情同意【Informed Consent】和私密性【Confidentiality】还是非常重要的。一般而言，访谈研究中的被访者会得到一封信，其中会解释研究目的(知情同意)，说明研究的赞助者(例如一个大学中的学院)、被访者需要在信上进行签名。和其他任何社会研究一样，访谈者必须许诺(并且要信守诺言)，所有被访者的姓名和标志性细节都会保密。

在研究者得到了伦理审查委员会(Institutional Review Board，IRB)的批准，他的问题也已经构想出来准备实施之后，他就需要寻找那些接受访谈的人——首先决定要访谈哪些类型的人(如性别、种族、年龄等)，以及访谈多少人(一般是 20～50 人)。然后，访谈者必须确定具体的被访者。抽样的一般问题在本卷书中其他地方已讨论过了；在这里只需要说的是，找人来访谈的困难部分地取决于访谈的主题。例如，如果研究是关于精神病人的，研究者就必须找到一家精神病医院来同意他作研究(在 2000 年代是非常困难的)，然后还要获得患者自己的同意——这两件事都是困难重重。研究与已婚男士相恋的单身女性遇到的问题可能会少一些，如果研究者是和被访者年龄相仿的女性就会更为容易。如理查森(Richardson，1985)所言，“找到其他女性来访谈并不困难”(p. x)，让她们讲自己的故事也不难，因为身处这种情境中的女性非常希望找人倾诉。

定性访谈是一种特殊的交谈，研究者会提出问题，被访者尽其所愿地讲述自己的故事(但还是有时间限制的！)；定性访谈持续

时间通常会在半小时到一个小时甚至更长。问题的类型包括概述式的("你能和我们谈谈……吗?")、追问式的("关于……,你能和我们再多谈一些吗?")、具体性的("当时你是怎么想的?")(Kvale, 1996, pp. 133-134)。访谈者不应该引导被访者,例如,他不应该问:"那件事是不是让你很难受?"而应该问:"你对那件事是什么感受?"问题的顺序应该是从最一般性的、最没有威胁性的问题开始,最后再问那些可能会让被访者感到不适的问题。被访者希望在访谈的前期和被访者建立起融洽关系(rapport),这样后面更为敏感的问题会更容易让人接受。

录音的定性访谈一般会由访谈者或其他人进行转录,形成访谈的文字稿【Transcription】。由访谈者来转录他自己做的访谈会更好一些,因为他最有资格解释其中一些含糊不清的陈述。付费(或不付费)的转录者缺乏访谈的切身经验,也没有动力细心认真地将口头录音转换为书面文字。在转录和分析访谈文字稿时,记住访谈是一种由轮流讲话形成的言语活动,这一点非常重要(Mishler, 1986)。

在定性访谈中,转录稿被用来作为分析的文本基础,正如田野日记在民族志【Ethnography】研究中的用处。对于访谈或者民族志文本的分析需要对转录稿进行反复阅读,确定其中的概念模式。在我对1950年代后期和1960年代早期的女性精神病患者及其丈夫的研究中,我发现女性从精神病院出院之后存在两种模式:一些女性及其丈夫想忘记这一切,否认她们曾经在精神病院待过;但是,访谈者每周的不断出现又让她们回想起了这一切。因此,她们要么试图结束访谈,要么无法结束访谈时(被访者有时极其听话!),她们努力地把访谈转变成社交活动,邀请访谈者留下来喝茶,或反过来开始向访谈者提问。其他女性则仍然很担心焦虑,想继续把访谈过程转变为一种治疗方式(Warren, 1987)。这里有一个例子:

Ann Rand 反复说,要是她不把访谈者请到家里来坐坐,她就再也不见访谈者了。访谈者对此闪烁其词,Ann 说:"我想你还是把我当成一个病人。在我眼里,你要么是朋友,要么是来管我的,现在你到底是哪一种?我就是这样看你的,你要么把我当朋友,要么把我当病人。"(p.261)

Oren 先生问我是否想和他们共进晚餐,我一直拒绝这样做,但他对此相当坚持。稍后他邀请我哪天下午去他们家,带个女人一起过来,放开好好玩一玩。……我强调说这是我的工作,大概希望让他这样来理解情况,不要重新界定。(p.261)

第一个例子中访谈者是女性,被访者也是女性:第二个例子中访谈者是男性,被访者也是男性,这并不是偶然的。1960年的家庭主妇不会想和一位男性访谈者交朋友——事实上,男性访谈者对女性被访者进行访谈是有困难的——同一时代的丈夫也不会邀请一位女性访谈者"去他们家,带个女人一起过来"。定性访谈的大量文献都说明了性别和其他个人特质对于访谈过程的重要性,因此这些因素对于作为分析对象的资料也有重要影响。1970年代和1980年代的一些女性主义定性访谈者,如安·奥克利(Ann Oakley),发现访谈方法特别适于引出女性的叙述(她们的日常生活常常是被人忽视的),特别适于打破社会科学研究中强势研究者和弱势研究对象之间存在的层级划分(Warren, 2002)。1990年代和2000年代的其他女性主义访谈者,如唐纳·勒夫(Donna Luff),发现女性内部在国籍、种族、社会阶级、政治意识形态上存在着深刻差异,这种差异并不一定能够用性别的同性来克服(Warren, 2002)。

回溯推理(representation)和后现代主义【Postmodernism】的议题对于社会科学中

定性访谈有很大的意义,正如它们对民族志和其他方法的意义一样。现代主义者(和新闻工作者)将被访者视为"回答载体(vessels of answers)",好像其回答就在那里等着访谈者来抽取一样,古布里乌姆和霍尔斯坦(Gubrium & Holstein,2002)对上述看法提出了挑战。古布里乌姆和霍尔斯坦以及其他人对访谈方法的批评(无论这些批评是来自于后现代主义者、女性主义者,还是持有其他立场的人)集中于访谈的认识论【Epistemology】:访谈是一种什么样的知识(knowing)?从中可以得到什么样的知识?过去十年间对上述问题涌现出了各种各样的回答,包括将访谈视为一种特殊形式的交谈(Mishler, 1986),从中只能知道语言形式的知识,别无其他。访谈是否只是一种言语活动,还是可以用来理解被访者在访谈之外的意义世界的资料,目前这还是一个可以讨论的问题。

虽然定性访谈是社会科学中最广泛使用的方法之一,但是它现在成为大量认识论争论的对象。如古布里乌姆和霍尔斯坦(Gubrium & Holstein,2002)指出的,我们现在生活在一个"访谈世界"中,每个人的观点都被认为是值得打探的,每个人都似乎想要被采访——如果不想被社会科学家访谈,他们至少想被媒体采访。定性访谈如同定量方法中的抽样调查访谈,有时可能被用来说明社会生活中那些既不能用定量方法也不能用民族志方法理解的方面。

定性访谈能够发现而实地研究不能发现的那些方面是什么?和定性访谈一样,民族志也在致力于理解社会成员的生活世界,但它的方式有所不同。回到精神病医院研究的例子中,实地研究者会观察当前医务工作者和患者的互动以及当前患者之间的互动,然后根据这些互动中观察到的模式和意义得出结论。一个访谈者则不仅探求当前的意义,而且探求与过去和将来有关的生平意义(biographical meaning);访谈的焦点不是互动(互动最好还是用观察方法来研究),而是被访者的陈述或故事。因此,对于实地研究者来说,ECT是医院用来制服那些不守规矩的病人的一种方便价廉的方法,而对于经历ECT的女性来说,这是一种对她们不能按文化规定的那样做一个好妻子和好母亲的惩罚。

在当今社会中,用访谈来进行社会科学研究既充满危险,同时又大有希望。在快速得到结果才最重要的访谈社会中,访谈是一种作研究的方便简易的方式,结果也产生了许多肤浅草率之作。但是,它也可以是理解人们生活和构建其日常生活和社会世界方式的一种宝贵工具。没有这样一种理解,优秀且有用的理论研究、定性研究或应用研究就无从谈起。无论社会科学家的目标是什么——构建或证明理论,进行大规模抽样调查研究,或提高社会状况——不理解生活在社会和世界中的人们的生活,这些目标都达不到。定性访谈与民族志携手为实现这样一种理解提供了工具。

——Carol A. B. Warren

(高勇译校)

参考文献

Gubrium, J. F., & Holstein, J. A. (2002). From the individual interview to the interview society. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research: Context and method* (pp. 3-32). Thousand Oaks, CA: Sage.

Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing*. Thousand Oaks, CA: Sage.

Mishler, E. G. (1986). *Research interviewing: Context and narrative*. Cambridge, MA: Harvard University Press.

Platt, J. (2002). The history of the interview. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of*

interview research: Context and method (pp. 33-54). Thousand Oaks, CA: Sage.

Richardson, L. (1985). *The new other woman: Contemporary single women in affairs with married men*. London: Collier-Macmillan.

Warren, C. A. B. (1987). *Madwives: Schizophrenic women in the 1950s*. New Brunswick, NJ: Rutgers University Press.

Warren, C. A. B. (2002). Qualitative interviewing. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research: Context and method* (pp. 83-101). Thousand Oaks, CA: Sage.

组内相关(Intraclass Correlation)

组内相关是一种对相关程度的测量,其基础为方差比率。严格地讲,它是一个比率,分子为测量对象的方差($\sigma^2_{对象}$),分母为测量对象的方差与误差项的方差之和:

$$\frac{\sigma^2_{对象}}{\sigma^2_{对象}+\sigma^2_{误差}}$$

它和被称为 Omega 方【Omega Squared, ω^2】的相关测量很相似,区别仅在于用来得出 $\sigma^2_{对象}$ 的样本估计的统计模型不同。组内相关是用随机效应模型得到估计量的(即样本中的对象是从潜在对象的总体中随机抽取出来的),而 ω^2 是用固定效应模型【Fixed Effects Model】得到估计量的(即样本中的对象就是我们感兴趣的对象,研究问题不改变,这些对象就不会改变)。

为了更好地理解,我们结合一个实际问题来介绍组内相关系数的应用:同胞兄弟姐妹在某一个生物测量特征上的相关程度有多大。如果用来得到相关估计的样本只包括一对一对的同胞(其中测量对象就是同胞),这时可以用皮尔逊积矩相关系数,但是用这种方法得出的估计量可能不是唯一的。举个例子可以说明其中的原因。有三个同胞对子{1,2},{3,4}和{5,6}。按原来顺序{1,2}{3,4}{5,6}得到的皮尔逊相关系数 r 和按同样的配对但不同的顺序{2,1}{3,4}{6,5}得到的皮尔逊相关系数 r 是不一样的,即使同胞数据完全一样。皮尔逊程序要求把对子中的成员完全武断地一个分到 X 组,一个分到 Y 组,但是事实上数据中的分组只是针对同胞对子的。此时如何得到一个同胞对子内的相关测量,而且还不用通过武断的分组来影响最后的估计值?早期的办法是把每个对子都再颠倒一次顺序,然后计算皮尔逊相关系数 r。在上面的例子中,这意味着对于以下对子计算皮尔逊相关系数 r:{1,2}{2,1}{3,4}{4,3}{5,6}{6,5}。

在同胞对子的例子中,把每对同胞的数据都加以颠倒提供了一种方便的解决方案,但是如果同胞数量扩展到包括 3、4 或者 k 个同胞时,这种方法就会越来越费劲。哈里斯(Harris,1913)提出了一种计算方法,费希尔(Fisher,1925)把他的方差分析法运用到了这个问题上。后来的方法表明,同胞相关是单因素随机效应方差模型的一种应用。在这个模型中,每个同胞的值都被概念化为三个独立部分的加总:要研究的生物测量特征的总体均值(μ);同胞效应(α),即同胞均值对于总体均值的偏差;同胞个体对于同胞均值的偏差(ε)。按其传统形式,这个模型可以写为 $Y_{ij}=\mu+\alpha_i+\varepsilon_{ij}$。此时,同胞相关就是 α 效应的方差除以 α 效应的方差与 ε 效应(误差效应)的方差之和:

$$\frac{\sigma^2_\alpha}{\sigma^2_\alpha+\sigma^2_\varepsilon}$$

换言之,它是这种特质的总方差中可以

归于不同同胞之间的差异的比例。对于那些在家庭内传递的特质来说，相关将会较高；对于那些不在家庭内传递的特质来说，相关将会较低。

将方差分析【Analysis of Variance】应用于同胞相关，解决了老方法在计算上的不便，可以很容易地推广到同胞数量不相等的数据分析中，澄清了相关作为一种方差比率的本质。它也提供了一种计算其他形式的组内相关的途径。同胞相关代表了使用单因素模型可以表示出来的相关。还有其他令人感兴趣的相关关系，会需要用到双因素或三因素模型。这些问题中许多涉及测量的信度【Reliability】：当 k 个评委对于 n 个选手逐个进行评分时，评委打分的信度如何？学校中学生在不同的老师培养下和在不同科目上的表现信度如何？上述问题都可用构建方差比率，即组内相关的方法来解决。事实上，两个最常见的信度测量，即 Cronbach α 系数【Cronbach's Alpha】和科恩 Kappa 系数(Cohen's kappa)，都可表示为组内相关。

——Kenneth O. McGraw

（高勇译校）

参考文献

Fisher, R. A. (1925). *Statistical methods for research workers* (1st ed.). Edinburgh, UK: Oliver & Boyd.

Harris, J. A. (1913). On the calculation of intraclass and interclass coefficients of correlation from class moments when the number of possible combinations is large. *Biometrika*, *9*, 446-472.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*, 30-46.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlation: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420-428.

编码者内部信度(Intracoder Reliability)

一个观察者或一个编码者对人的行为或者其他现象(如电视节目中的暴力行为、儿童书籍中暗含的性别刻板印象等)的判断常常会在不同时间中有起伏浮动。编码者(评判者或观察者)内部信度估计的就是同一个编码者在不同时点的判断的相对一致性。具体而言，它估计的是由于如细心程度、情绪、干扰、疲倦、一段时间内目标对象的波动等因素造成的不一致数量(即测量误差【Measurement Error】)。编码者内部信度不同于评判者信度【Interrater Reliability】，后者评估的是两个或多个评判者(或编码者)的判断的相对一致性。

估计编码者内部信度的传统程序类似于再测信度【Test-Retest Reliability】。一个编码者在有一定间隔的两个不同时间场合(occasions)对于目标行为(如电视中暴力场景的数量)进行判断。然后用对称测量【Symmetric Measures】中讨论的各种公式对这两组判断进行相关分析。相关系数就代表了编码者内部信度的估计。

在下面的篇幅中，我们将介绍一种基于概化理论【Generalizability】或 G 理论对多个时点中的编码者内部信度进行评估的方法，它更为灵活，但较少有人使用(Brennan, 1983; Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Shavelson & Webb, 1991)。假定一个编码者在连续四天中每天都观察 10 个自闭症儿童在 1 小时内的眼神交流次数。观察记录如表 1 所示。

表 1 一个编码者观察到的 10 个自闭症儿童的眼神交流数量

	时点				儿童	
	第 1 天	第 2 天	第 3 天	第 4 天	均 值	标准差
儿童 A	0	2	2	0	1	1.15
儿童 B	2	3	1	1	1.75	0.96
儿童 C	2	0	1	2	1.25	0.96
儿童 D	1	3	0	0	1	1.41
儿童 E	1	0	0	1	0.5	0.58
儿童 F	4	3	2	2	2.75	0.96
儿童 G	1	0	3	0	1	1.41
儿童 H	2	1	0	3	1.5	1.29
儿童 I	3	1	2	0	1.5	1.29
儿童 J	3	1	0	0	1	1.41
均值	1.9	1.4	1.1	0.9	总均值	1.33
标准差	1.2	1.26	1.1	1.1	总标准差	1.19

基于表 1 中的内容,我们可以确定编码者的观察存在变异的四个来源:(a)自闭症儿童之间的差异(即儿童的主效应【Main Effect】);(b)同一编码者在四个时间场合上的差异(即时间场合的主效应);(c)不同时间场合中儿童有着不同表现(儿童与时间场合的交互效应【Interaction Effect】);(d)未知误差(如随机误差【Random Error】和未知的系统误差【Systematic Error】)。所有的误差来源都被假定是随机的,以便我们能够回答如下问题:"编码者在一个时间场合的观察在多大程度上能够推广到其他时间场合?"由于从数据中不能分辨第三和第四个来源,这两者合起来被认为是残差【Residual】。源自上述来源的变异程度可用方差成分模型【Variance Components Model】来进行估计。

基于表 2 中的公式,可以估计出方差成分,表 3 中显示了这些方差成分。可以看出,儿童的方差成分仅仅解释了总方差的 4%,这表明基于一个时间场合的观察并不能区分儿童的行为。残差占据了方差的一大部分比例,这表明儿童的眼神交流在不同时间场合有着很大波动。

表 2　方差来源、相应的符号和计算公式

方差来源	平方和	自由度	均　方	估计出的方差成分
儿童(p)	SS_p	$df_p=n_p-1$	$MS_p=SS_p/df_p$	$\hat{\sigma}_p^2=(MS_p-\hat{\sigma}_{po,e}^2)/n_o$
时间场合(o)	SS_o	$df_o=n_o-1$	$MS_o=SS_o/df_p$	$\hat{\sigma}_o^2=(MS_o-\hat{\sigma}_{po,e}^2)/n_p$
残差(po,e)	$SS_{po,e}$	$df_{po,e}=df_p \times df_o$	$MS_p=SS_{po,e}/df_{po,e}$	$\hat{\sigma}_{po,e}^2=MS_{po,e}$

注：n_p 是儿童数目，n_o 是时间场合数目。

表 3　基于表 1 中数据的方差来源

方差来源	平方和	自由度	均　方	估计出的方差成分	占总方差的百分比/%
儿童(p)	13.53	9	1.50	0.046	4
时间场合(o)	5.68	3	1.89	0.057	4
残差(po,e)	35.58	27	1.32	1.318	93

为了评估编码者内部信度，我们首先要确定感兴趣的是何种测量误差，这要由研究者的阐释和决策来决定。研究者可以基于一个儿童与其他儿童相比表现如何(即相对决策)来解释数据，也可以基于眼神交流的绝对数量(即绝对决策)来解释数据。相对决策的测量误差(σ_R^2)可以用下式估计：

$$\hat{\sigma}_R^2=\frac{\hat{\sigma}_{po,e}^2}{n_o^*}$$

而绝对决策的测量误差(σ_A^2)可以用下式估计：

$$\hat{\sigma}_A^2=\frac{\hat{\sigma}_o^2}{n_o^*}+\frac{\hat{\sigma}_{po,e}^2}{n_o^*}$$

其中的 n_o^* 是研究中(表 1)用到的时间场合数目，或者研究者选择用来提高观察一致性的时间场合数目。

对于相对决策来说，编码者内部信度(或者在 G 理论中被称为概化系数)以下式估计：

$$G_R=\frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2+\hat{\sigma}_R^2}$$

对于绝对决策来说，编码者内部信度以下式估计：

$$G_A=\frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2+\hat{\sigma}_A^2}$$

表 4 中显示了在不同决策类型和不同数目的时间场合下对编码者内部信度的估计。如表中所示，一个编码者研究每个儿童的次数越多，测量误差就越小。编码者的观察至少需要 80 天，才能达到令人满意的一致性水平。数据表明，如果编码者只观察儿童 4 天时，编码者内部信度是很差的。

表 4　在不同决策类型和不同数目的时间场合下对测量误差和编码者内部信度的估计

时间场合	相对决策		绝对决策	
	测量误差	编码者内部信度	测量误差	编码者内部信度
1 天	1.318	0.033	1.364	0.032
2 天	0.659	0.065	0.682	0.063
3 天	0.439	0.094	0.454	0.091
4 天	0.329	0.122	0.341	0.118
5 天	0.263	0.148	0.272	0.144

续表

时间场合	相对决策		绝对决策	
	测量误差	编码者内部信度	测量误差	编码者内部信度
10 天	0.131	0.258	0.136	0.252
20 天	0.065	0.411	0.068	0.402
30 天	0.043	0.511	0.045	0.502
40 天	0.032	0.582	0.034	0.574
80 天	0.016	0.736	0.017	0.729
120 天	0.010	0.807	0.011	0.801
160 天	0.008	0.848	0.008	0.843

——Peter Y. Chen
Autumn D. Krauss
（高勇译校）

参考文献

Brennan, R. L. (1983). *Elements of generalizability*. Iowa City, IA: American College Testing Program.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability of scores and profiles*. New York: Wiley.

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.

调查者效应(Investigator Effects)

调查者效应是科学研究中源于调查者的那些假象来源和误差来源。科学家对于研究结果通常是无意识的，主要有两种影响：第一种类型的影响可以说发生在科学家的眼、手、脑中。它的发生并不会影响研究中参与者或者动物被试的实际反应。第二种类型的调查者效应是互动性的。它通过影响研究对象的实际反应而起作用。

非互动性的效应包括观察者效应、诠释者效应、意图性效应。

(1)观察者效应(observer effect)。是指科学家在感知或记录所调查事项时的观察误差。对涉及 314 位观察者、总共139 000次观察记录的 21 项研究的分析表明，大约1%的观察是存在误差的，误差发生时 2/3 的情况都是误差方向有利于观察者的假设。

(2)诠释者效应(interpreter effect)。是指不同的科学家会对同样的观察给出不同的理论诠释。例如，多年来调查者对现有的精神疗法有效性研究的诠释有重大分歧。直到对数百项精神疗法有效性研究进行了系统的定量汇总之后，这一诠释者效应的具体议题才算得以圆满解决。

(3)意图性效应(intentional effect)。是指科学中不诚实的情况。也许最常见的例子就是编造数据，不对要研究的事件进行诚实的观察，而是编造事实上并未发生而研究者希望得到的观察。

交互性的调查者效应包括生物社会效应、心理社会效应、情境效应、示范效应和期望效应。

(1)生物社会效应(biosocial effect)。是指由于调查者的性别、年龄、民族等因素，研究参与者的反应可能出现的差异。

(2)心理社会效应(psychosocial effect)。是指由于调查者的人格或社会地位等因素，研究参与者的反应可能出现的差异。

(3)情境效应(situational effect)。是指由于调查者在情境变量上的差异，如调查者的研究经历、与参与者的事先熟悉程度、从先前接触过的参与者中得到的反应等，研究参与者的反应可能出现的差异。

(4)示范效应(modeling effect)。是指

由于调查者自身对所执行任务的反应有所差异,研究参与者的反应可能出现的差异。

(5)期望效应(expectancy effect)。是指由于调查者期望从参与者那里得到的反应类型有所差异,以及研究参与者的反应可能出现的差异。在社会科学或行为科学中,调查者持有的期望或假设已经被证明会影响调查者的行为,最终会导致参与者给出调查者期望的反应。这种效应常被称为实验者期望效应【Experimenter Expectancy Effect】,这是已经被发现发生在实验室之外的情境中的人际期望效应(interpersonal expectancy effects)的一种特例。

——Robert Rosenthal

(高勇译校)

*也可参见皮格马利翁效应【Pygmalion Effect】。

参考文献

Rosenthal, R. (1976). *Experimenter effects in behavioral research: Enlarged edition*. New York: Irvington.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart and Winston.

Rosnow, R. L., & Rosenthal, R. (1997). *People studying people: Artifacts and ethics in behavioral research*. New York: W. H. Freeman.

原生编码(In Vivo Coding)

要理解原生编码,先要理解编码【Coding】。编码就是一个概念或一个词,它表明了"这项资料中说的是什么事"。另一方面,编码也是一个分析过程,在这一过程中要逐行或逐段地查阅资料(根据风格而定)来找到重要的事件、经历或感受等,然后再用概念来指代这些内容(Strauss & Corbin, 1998)。有时,分析者命名这些概念。另外一些时候,被访者的描述用语是如此贴切,因此这些用语成了最后被选定的概念。被访者提出的这些描述用语就被称为"原生概念"(Glaser & Strauss, 1967)。

原生概念(vivo concepts)通常是非常能说明问题的、能揭示本质的精辟词语。原生编码中有意思的是,在被访者表达出这一想法的当下,研究者就知道这就是要注意的东西。这个用语比分析者能想出的任何词汇都更准确地表达出了其中的意义(Strauss, 1987)。例如,假定一个研究者正在研究鸡尾酒会上的事情。一个被访者说了下面一段话:

好多人讨厌去鸡尾酒会,可我喜欢去。我去鸡尾酒会的时候,我利用这个机会来"现场办公"。这些看起来是些闲言碎语,但其实我脑子里会盘算以后的业务往来、我想约会的女人、可能的网球搭档甚至投资机会。

"现场办公(to work the scene)"是个很棒的概念。它没有告诉我们鸡尾酒会上的任何具体事情,但是它确实表达出了这个人在干的事情。如果他在这么干,可能其他人也在"现场办公"。研究者继续访谈时可以把这个概念记在脑子里,看看其他人描述的事情是否也可以归类成"现场办公"。

与此不同的是,源于分析者的编码会是下面这样的。研究者观察另一个鸡尾酒会时,注意到两个女性在兴奋地讲着她们从事的工作项目。在另一块地方,一个男性和一个女性也在谈论他们的工作,除此之外,另一群人也在做类似的事情。研究者把这些鸡尾酒会上关于工作的谈话称为"社交/工作交谈"。这个词描述了发生的事情,但它

不如“现场办公”精辟或有趣。

不是每个访谈或观察都能够得出有趣的原生编码，但是如果资料中出现了原生编码，研究者应当好好利用它。重要的是，分析者要对资料内容保持机敏和敏感，被访者的用语往往是最佳的表达方式。编码是个费力而烦琐的过程，特别是进行逐行分析时。但是，正是通过这些烦琐的编码才能获得那些财富，那些被我们称为“原生编码”的用语。

——Juliet M. Corbin

（高勇译校）

参考文献

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory*. Chicago: Aldine.

Strauss, A. (1987). *Qualitative analysis for social scientists*. Cambridge, UK: Cambridge University Press.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative analysis*. Thousand Oaks, CA: Sage.

同构(Isomorph)

同构就是与另一种理论、模型或结构类似、等价甚至相同的理论、模型或结构。例如，两个理论可能在概念上是同构的，但是使用不同的经验测量，因此在经验意义上是有区别的。或者，两个理论可能在概念上或者经验上是不同的，但是在其预测上却是同构的。

——Michael S. Lewis-Beck

（高勇译校）

项目反应理论(Item Response Theory)

定义与应用领域

测试和问卷都包括一些题项（用 J 表示），每个题项用来测量同一种潜在心理能力、人格特质或态度的一个方面。项目反应理论（IRT）模型使用 N 个答题人构成的样本中 J 个题项收集的资料，来构建测量此种能力或特质的量表。

题项 j ($j=1, \cdots, J$) 上的得分可以用一个随机变量 X 来表示，其现实值（realizations）为 x_j。题项得分可以是：

- 二分的，表明对这一题项的回答是正确（$x_j=1$）还是错误（$x_j=0$）。
- 定序的，表明答题人同意一个具体陈述的程度（有序整数值，$x_j=0, \cdots, m$）。
- 定类的，表明答题人选择的具体答案类别，如在多项选择题中，其中一个选项正确，其他选项不正确，因此是定类测量层次。
- 连续的，如解决问题所花费的时间。

从数据中可以估计出 J 个题项的特性，如它们的难度。然后利用这些特性可以决定选择哪些题项进入笔试或问卷中，或者更先进的电脑化的测量程序中。因此题项特性在测量构建中具有技术作用，有助于产生高质量的量表来对个体进行测量。

在 IRT 背景中，潜在于题项成绩背后的能力、人格特质、态度被称为潜特质。一个潜特质可以用随机变量 θ 表示，每个参加用来测量 θ 的测试的人 i ($i=1, \cdots, N$) 都有一个量表值 θ_i，IRT 的主要目标是从每个人在 J 个可观察的题项得分中估计出其 θ 值。这些估计出的测量值可以用于人际比较或

与外在行为标准比较。此类比较构成了对个体进行决策的基础。

IRT 源起于 1950 年代和 1960 年代(如 Birnbaum, 1968),随后达到了繁荣。重要的应用领域包括:

- 教育测量,其中测试被用于给学生考试打分,选择学生进行补救性教育以及进入高一级教育(如升入大学),对专业工作者的技能进行资格考察。
- 心理学,其中用智商测量来诊断儿童的认知能力,以解决其在学校中的学习和注意力问题,用人格量表来选择病人进入临床处理,用能力检试来选择工厂和政府中的工作和岗位。
- 社会学,其中会测量对堕胎或者单亲家庭抚育的态度,也涉及如疏离、马基雅维利主义、宗教性等潜特质。
- 政治学,其中会用问题来测量选民对政治家和政党的偏好,以及政治效能感和对政府环境政策的看法。
- 医学研究,其中会测量患者与健康相关的生活质量,这些患者可能遭受了引发持久性生理损伤的事故,或者做了大手术,或者接受了实验性药物的长期治疗,或者接受了严重影响患者日常生活经历的其他治疗。
- 市场研究,其中会测量消费者对产品和品牌的偏好。

上述应用中都需要一个量化的量表来测量人们的能力,而这正是 IRT 所能做的。

本词条将首先介绍 IRT 模型的常用假定。然后,再进行几项区分来对不同的 IRT 模型进行分类。最后,讨论几个有用的 IRT 模型应用。

IRT 模型的常用假定

测量的维度性

第一个假定列举出汇总测试或问卷参加者的成绩所需要的参数数目。例如,如果一个测试是用来测量空间方位感的(有些智商测试中会涉及这些题项),那么就会用假定只有一个参数的 IRT 模型来描述数据。这个参数代表了每个人的空间方位感水平。只有一个参数的 IRT 模型是严格的单维度(unidimensional, UD)模型。另一个测试中的题项可能测量的是算术和词语理解能力的混合,例如,算术练习以短故事的形式出现。这时,就需要一个两维度的 IRT 模型。其他测试成绩可以更加复杂,需要用到多维度的 IRT 模型来解释数据结构。例如,在数学考试中,一些题项需要了解商店及其出售的商品(如计算收银台应该找的钱数),另一些题项需要地理知识(如计算城市间的距离)。从本质上讲,单维度模型假定测试中的所有题项测量的是一个主导性潜特性以及许多扰动特性,只要测试足够长,这些扰动特性就不会干扰到对主导特性 θ 的测量。例如,一个内倾性人格量表也可能会测量到焦虑、社会智商、语言理解,但这些内容都只是在一两个题项中略有涉及,内倾性才是回答的主导驱动力。

题项之间的关系

多数 IRT 模型假定,如果已经了解某人在潜特质上的位置,同时也知道了 J 个题项的边缘频数分布,那么就可以重构出这个人各个题项得分的联合分布。定义向量 $\mathbf{X}=(X_1,\cdots,X_J)$,其现实值为 $\mathbf{x}=(x_1,\cdots,x_J)$,$\boldsymbol{\theta}$ 为解释答题人测量成绩所需的潜特质向量。在 IRT 中,边缘独立性特性被称为局部独立【Local Independence, LI】,定义如下:

$$P(\mathbf{X}=\mathbf{x}\mid\boldsymbol{\theta})=\prod_{j=1}^{J}P(X_j=x_j\mid\boldsymbol{\theta}) \qquad (1)$$

在行为层面,LI 意味着在测试实施中,每道题项的回答都独立于其他题项。在回答问题时,不会由于练习而学到或了解到后面的题项答案;因此,测量程序不会影响测量结果。但是,通过练习来提升学生能力可能是测试的目标,这在动态测试中很常见。

此时,IRT 模型就会假定练习会接连产生更多的技能,这些技能可以用新的潜特质表示。这意味着,测试时不能用 **θ** 解释的题项之间的相依可以通过加入更多的 θ 来解决。

题项与潜特性之间的关系

对于单维度二分题项来说,题项反应函数(item response function,IRF)描述了题项得分 X_j 与潜特质之间 θ 的关系,表达式为 $P_j(\theta)=P(X_j=1|\theta)$。图 1 显示了两个典型的单调递增的 IRF(称为 M 假定),这是来自下面将要介绍的三参数 Logistic 模型。M 假定的意思是,θ 越高,回答人对题项回答的正确概率就越高。但是这两个项目反映了函数在三个方面有所差异。

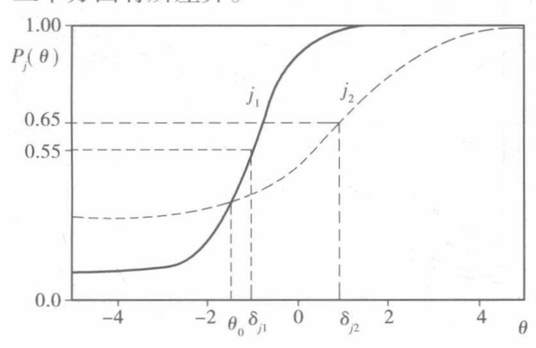


图 1　三参数 Logistic **模型下的两个** IRF

(**参数值**:$\gamma_{j1}=0.10,\gamma_{j2}=0.30;\delta_{j1}=-1.00,\delta_{j2}=1.00;\alpha_{j1}=2.00,\alpha_{j2}=1.00$);**交叉点的位置为** $\theta_0=1.46;P(\theta_0)=0.36$

首先,IRF 的斜率不同。IRF 越陡峭,代表着题目与潜特性之间的关系越强。IRT 模型中有一个参数与 IRF 的最陡峭斜率有关。题项 j 的这个斜率参数表示为 α_j,它可以类比于 Logistic 回归【Logistic Regression】模型中的回归系数【Regression Coefficient】。题项 j_1(实线)的斜率比题项 j_2(虚线)更陡,因此,它与 θ 的关系更强,能够更好地区分 θ 的高分值和低分值。

其次,题项反应函数的位置不同。每个项目反应函数的位置取决于参数 δ_j,它决定了 IRF 位于最高条件概率和最低条件概率值的中间点上的 θ 值。题项 j_1(实线)的位置比题项 j_2(虚线)更靠左,它们的位置参数表明了这一点:$\delta_{j1}<\delta_{j2}$。由于它们的斜率不同,两条 IRF 交叉了。结果,对于交叉点左边的那些 θ 而言,题项 j_1 比题项 j_2 更不可能正确回答。因此,对于这些 θ 而言,题项 j_1 比题项 j_2 更难。对于 θ_0 右边的那些 θ 而言,题项难度的顺序正好相反。

最后,IRF 的低端渐进线(用 γ_j 表示)不同。题项参数 γ_j 是那些 θ 值极低的人能够正确回答的概率。在图 1 中,题项 j_1 有更低的 γ 参数。这个参数对多项选择题特别重要,因为那些 θ 值较低的人常常能够猜测到正确答案的概率并不是 0。因此,多项选择题项 j_2 比题项 j_1 更容易被猜中。

三参数 Logistic 模型(three-parameter logistic model)。图 1 中的 IRF 可以用三参数 Logistic 模型来定义。这个模型是基于 UD 和 LI 假定基础上的,用一个 Logistic 函数来定义题项反应函数,函数中包括了前面讨论的三个题项参数:

$$P_j(\theta)=\gamma_j+(1-\gamma_j)\frac{\exp[\alpha_j(\theta-\delta_j)]}{1+\exp[\alpha_j(\theta-\delta_j)]} \quad (2)$$

单参数 Logistic 模型(one-parameter logistic model)。另一种众所周知的 IRT 模型是单参数 Logistic 模型或者 Rasch 模型,它也是基于 UD 和 LI 假定基础上的,它假定对于测试中的所有题项 J 来说,$\gamma=0$ 且 $\alpha=1$(取值为 1 是主观判定的;这意味着对所有 j 来说,α_j 都是一个常数 a)。因此,这个模型:(a)不适用于拟合多项选择项数据,因为那种情况下 γ 取值为正;(b)假定所有题项得分与潜特质之间的关系强度相同(对于所有 j 来说,$\alpha_j=a$);(c)题项之间只是在难度 $\delta_j,j=1,\cdots,J$ 上有所差异。

线性 Logistic 多维模型(linear logistic multidimensional model)。图 2 显示了一个题项反应函数,它是一个三维平面,反应概

率取决于两个潜特质，$\boldsymbol{\theta}=(\theta_1,\theta_2)$。在 θ_2 方向上比在 θ_1 方向上更加陡峭。这意味着题项回答正确的概率更取决于 θ_2，而不是 θ_1。因为 θ_2 比 θ_1 对于正确回答更重要，使用这一测量题项，人们在 θ_2 维度上要比 θ_1 维度上能够得到更好的鉴别。其他用来测量综合特质 $\boldsymbol{\theta}$ 的题项的两个斜率可能会有所不同。这个题项的位置（图 2 中看不到）与空间原点与沿原点方向的最陡斜率之间的距离相关（但不等同）。γ_j 参数代表了在两个 θ 都很低时答案正确的概率。

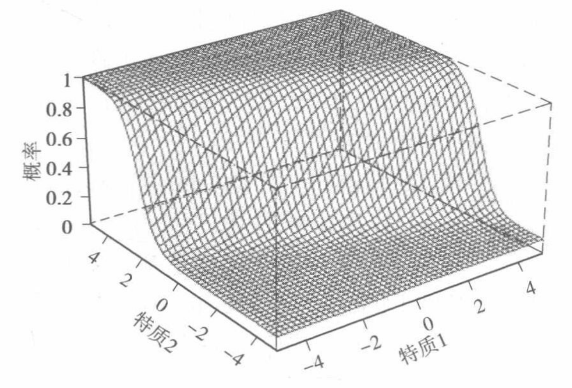


图 2　线性 Logistic 多维模型中的题项反应平面
（参数值：$\gamma_j=0.10$；$\delta_j=0.00$；$\alpha_{j1}=1.00$；$\alpha_{j2}=2.50$）

图 2 中的题项反应平面源于 Reckase 的线性 Logistic 多维模型（Van der Linden & Hambleton, 1997, pp. 271-286）。这个 IRT 模型中的 $\boldsymbol{\theta}$ 是多维的，斜率参数用一个向量 $\boldsymbol{\alpha}$ 表示。题项反应平面的表达式如下：

$$P_j(X_j=1\mid\boldsymbol{\theta})=\gamma_j+(1-\gamma_j)\frac{\exp(\boldsymbol{\alpha}_j'\boldsymbol{\theta}+\delta_j)}{1+\exp(\boldsymbol{\alpha}_j'\boldsymbol{\theta}+\delta_j)}\qquad(3)$$

等级反应模型（graded response model）。最后，图 3 的实线显示了 Samejima 的等级反应模型中一个题项的题项梯级反应函数（item step response function，ISRF），此时题项得分是分多个级别的。对于有序整数题项得分（$x_j=0,\cdots,m$）和单维潜特质来说，每个条件反应概率 $P(X_j\geqslant x_j\mid\theta)$ 都分别用 Logistic 函数来建模。对于同一题项的不同 ISRF，这些函数的位置参数有所不同，但是斜率参数是相同的。对于不同的题项，这些函数的斜率参数有所不同。可以看出，对于 $x_j=0$ 来说，ISRF 等于 1；对于某一固定题项来说，其他 m 个 ISRF 都必然不可能交叉。但是不同题项的 ISRF 是可以交叉的，因为斜率参数不同（参见图 3）。多级项目反应理论比二分项目反应理论在数学上更为复杂，因为它涉及更多的反应函数，项目参数也更多。除此之外，还有许多种其他项目反应理论模型（具体可以参看 Van der Linden & Hambleton, 1997）。

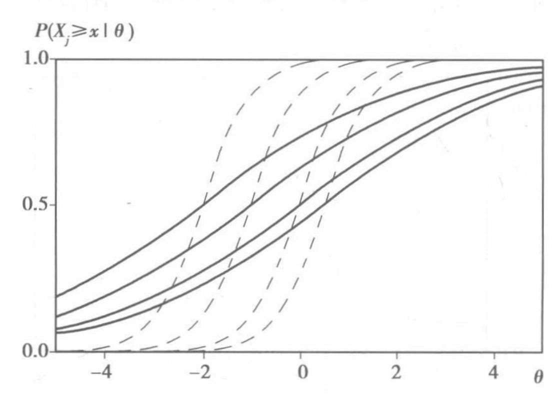


图 3　等级反应模型下两个题项的 ISRF，
各有 5 个回答类型（未给出参数值）

非参数模型（nonparametric models）。前面讨论的模型都是参数模型，因为其 IRF 和 ISRF 都是 θ 的参数函数。非参数项目反应理论模型只是对 IRF 施加有序约束，而放宽了更加严格的参数定义。这样做的目的是对测量模型不再施加不恰当的测试数据结构，但是仍然施加足够的结构来产生人们在 θ 上的定序测量。这种排序是通过题项得分的加总得到的，它将 θ 替换为测试成绩的汇总。因此，获得灵活性的代价是损失了参数函数（特别是 Logistic 函数）的数学特征的方便性（对参数 IRT 模型和非参数 IRT 模型的讨论，请参见 Boomsma, Van Duijn, & Snijders, 2001；对非参数 IRT 模型的介绍，请

参见 Sijtsma & Molenaar,2002)。

项目反应理论模型的应用

测试的建构和校准(test construction, calibration)。IRT 模型就是一种测量量表,无论是定序的还是定距的。如果 IRT 模型能够拟合数据,这种量表就被认为在这个具体测试中成立。量表被加以校准和转换,以便进行方便的解释。例如,θ 量度可以转换为众所周知的总成绩量表或者百分制量表。这种量表构成了个体诊断的基础(如在教育和心理学测量中),也是科学研究的基础。

等值、题库和自适应测验(equating and item banks, adaptive testing)。θ 量度可以很方便地将对相同潜特质的不同测试成绩进行量纲等值化,其目的是直接比较参与不同测试的学生成绩。等值可用于构建题库,题库由测量相同潜特质,但难度和其他特性有所不同的几百道题项组成。新的测试可以从题库中抽取组合。对个体的测试可以从题库中一道一道地抽取,以便尽快降低估计出 θ 的测量误差。这被称为自适应测验。在教育、工作选择和岗位选择中,它可以很方便地用于大规模测验项目。

差别性题项运作(differential item functioning)。不同背景的人常常会面临相同的测量工具。一个重要的问题是,有着相同 θ 水平但是不同背景(如性别、社会经济背景、民族)的人对测试题项是否会有相同的反应概率。如果反应概率并不相同,这个测试就被称为在题项运作展现出了差别性。这可以用 IRT 方法来进行调查。在群体间功能有差异的题项应当用功能无差异的题项替换。

个体吻合分析(person-fit analysis)。答题人可能会对题项的形式迷惑不解;他们可能害怕这种被评估的测试情景;他们可能会低估测试水平,忽视有些问题的深度。这些机制以及其他机制可能会产生与 IRT 模型预测不相符的 J 个题项的得分模式。个体吻合方法可以找到那些不相吻合的题项得分模式。它们有助于诊断出引发这种异常模式的行为。

认知性 IRT 模型(cognitive IRT model)。最后,认知建模已经不再认为测量仅仅是把分值赋给人,因为产生这些分值的认知过程或者解题策略也是 IRT 模型的部分,测量是与心理学解释相关联的。这种方法使我们能够找到那些未被充分掌握的技能,或者在理论层面上更好地理解测试成绩背后的过程。

——Klaas Sijtsma

(高勇译校)

参考文献

Birnbaum, A. L. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Boomsma, A., Van Duijn, M. A. J., & Snijders, T. A. B. (Eds.). (2001). *Essays on item response theory*. New York: Springer.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.

Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory*. Thousand Oaks, CA: Sage.

Van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory*. New York: Springer.

J

刀切法(Jackknife Method)

刀切法是一种再抽样程序,可以用于产生一个经验抽样分布【Sampling Distribution】。此类分布可以用于进行统计假设检验、构建置信区间【Confidence Interval】、估计标准误【Standard Error】。刀切法的作用类似于自助法【Bootstrap】和排列法(或者随机选择法)。

罗杰斯(Rodgers,1999)提出了一种分类法【Taxonomy】,根据再抽样出来的样本量大小和抽样是有放回还是无回放来对这些方法进行分类。刀切法需要无放回地进行重复抽样,抽出样本比原始样本要小。刀切法最初被定义为一种将一个或多个数据点随机删除的方法,这在概念上等同于无放回的随机抽样。

刀切法最初是由奎诺依勒(Quenouille,1949)和图基(Tukey,1958)提出。布拉德利·埃弗龙(Efron,1982)将其命名为"Quenouille-Tukey 刀切法"。莫斯迪勒和图基(Mosteller & Tukey,1977)将这种方法概括为:"基本思想是评估构成数据的各组分数据的效应……方式是评估删除此组数据后得到的数据集的后果"(p.133)。刀切法已经被加以扩展,如罗杰斯(Rodgers,1999)所言:"定义观察数据为一个得分总体。从这个总体中无放回地进行随机抽样,以使各组中的得分数量都比原始样本中更少"(p.46)。总而言之,经验抽样分布是通过删除一个、多个或者一组观察数据而得到的;然后用新的样本计算相关的统计量;通过多次重复上述过程得到这些统计量的分布。

举例来说,考虑一个将10个参与者随机分配【Random Assignment】到治疗组和对照组的研究设计(每组中 $n=5$)。这种治疗方法是否有可靠的效果呢?在组间无差异的虚无假设【Null Hypothesis】下,10个得分被分到这组还是那组,完全取决于机会。对于虚无假设通常的正规检验需要计算一个彼此独立的两组数据的 t 统计量,它测量了两组之间的标准化均值差异,然后在假定虚无假设成立时的理论 t 分布下进行评估。这一评估基于许多参数统计假设,包括正态性、误差独立性、方差同质性。我们也可以换个方法,虚无假设可以定义为使用再抽样得到的经验 t 分布,这一过程就需要更少的假设。

用刀切法,我们可以从每组中删除1个观察(或者,这相当于从原始的10个观察中进行无放回的抽样,直到每组都有4个观察)。共有3 150个可能的再抽样。在每个新样本中都计算 t 统计量,可以得到 t 统计量的分布,这接近于在虚无假设下总体的 t 分布。然后再与这个经验抽样分布相联系,来评估最初得到的 t 统计量,以便对于组间差异作出结论。进而,我们可以用再抽样的 t 统计量的标准差估计总体分布的 t 统计量

的标准误，并构建相应的置信区间。

——Joseph Lee Rodgers

（高勇译校）

参考文献

Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans*. Philadelphia: Society for Industrial and Applied Mathematics.

Mosteller, F., & Tukey, J. W. (1977). *Data analysis and regression: A second course in statistics*. Reading, MA: Addison-Wesley.

Quenouille, M. (1949). Approximate tests of correlation in time series. *Journal of the Royal Statistical Society, Series B*, *11*, 18-84.

Rodgers, J. L. (1999). The bootstrap, the jackknife, and the randomization test: A sampling taxonomy. *Multivariate Behavioral Research*, *34*, 441-456.

Tukey, J. W. (1958). Bias and confidence in not quite large samples [abstract]. *Annals of Mathematical Statistics*, *29*, 614.

K

关键知情人(Key Informant)

关键知情人也被称为关键咨询人能就某种文化、某个社会或某种社会场景的诸要素给予全面而深入的专门信息。此类深度信息通常是在自然的非正式场合中(如关键知情人的家中)通过重复进行几次半结构化访谈【Semi-Structured Interview】或非结构访谈【Unstructured Interview】得来的。除了信息的深度和广度,关键知情人区别于其他被访者之处还在于他和研究者之间的关系是持久而紧密的。社会科学中主要包括三种类型的被访者:被试(subjects)、回答人(respondents)和知情人。被试通常是在社会实验中被访谈的个体。回答人是在社会调查【Survey】中被访谈的个体。知情人通常是在实地研究场景中以更深度、更少结构化的方式(如半结构化或非结构访谈)被访谈的个体。在任何一项研究中,个体的角色可以在知情人、回答人、被试之间进行转换,这取决于访谈情境(如调查、深度访谈【In-Depth Interview】)。关键知情人是知情人中的一种类型,其中的访谈者(如民族志学者)和被访者之间存在着特殊关系。

选择关键知情人的基础是知情人的特征、知识与研究者的互信关系(rapport)。约翰逊(Johnson,1990)描述了选择知情人的两组主要标准。第一组标准与专家知识的形式和类型以及知情人的理论代表性(theoretical representativeness)有关,在此理论代表性指从理论上的重要因素(如地位、组织中的位置)和资料呈现特征(如收入分布中知情人的位置)方面评估的被访者特征。第二组标准关切的是知情人的人格、人际技巧、研究者与知情人的默契。第一组标准在研究目标框架下评估知情人,第二组标准反映了研究者和关键知情人之间存在合作关系的重要本质。研究者理解了个体在理论标准领域内的角色后,就应基于互信关系的潜在可能来选择关键知情人,这降低了单纯基于个人特征进行选择可能产生的偏差。

关键知情人与其他几种被访者的区别还在于他们在研究工作不同阶段的涉足程度有所不同。关键知情人及其专业知识和建议不仅是资料的来源,而且可以用于发展出更为有效的正式调查工具(如对调查问题进行更好的表述),对文化事件或场景不断提供评论,帮助解释系统收集到的资料(如抽样调查结果)中显示出来的研究结果,帮助确定某些资料(如家谱)的效度【Validity】和信度【Reliability】。关键知情人是任何民族志研究的重要组成部分。

——Jeffrey C. Johnson

(高勇译校)

参考文献

Johnson, J. C. (1990). *Selecting ethnographic informants*. Newbury Park, CA: Sage.

Werner, O., & Schoepfle, G. M. (1987). *Systematic fieldwork* (2 vols.). Newbury Park, CA: Sage.

Kish 表(Kish Grid)

Kish 表是 1949 年由莱斯莉·基什(Leslie Kish)提出的,通常在大规模抽样调查【Survey】中使用。在等概率抽样【Sampling】中,访员进入样本地址或样本家庭时可能发现有多个案例符合调查条件,这时就可以用 Kish 表技术来随机抽取案例。例如,从某一地区的所有邮政地址清单中抽取地址样本,样本设计假定清单中的每个地址只包括一个家庭。访员进入样本地址时发现这个地址中居住着多个家庭,就需要随机选择一个家庭(或者样本设计中设定的数目)来进入样本。更重要的是,这一程序必须是系统的,要确保每个家庭被选择的概率是已知的。

户中抽人同样存在这一问题。一户中可能只有一人或者有多人。包括多人的家庭就成了总体元素的一个集合,必须把它们从家庭样本变成个人样本。如果调查设计要求被访谈的个体必须是某种特殊类型的(如必须是户主或者包括所有 16 周岁以上的人),那就不存在选择问题了。但是,如果调查设计要求从每户中选择一个人,就可以用 Kish 表从每个选中的家庭中随机选择进入样本的个人。

Kish 表是一个简单的程序,访员很容易在实地进行操作。它提供了一种清晰一致的程序来随机选择案例进入样本,而且选择概率是已知的。在进行选择之前,访员首先需要用一种客观的方式来对家庭中的成员或者居住者进行排序。为了从家庭中选出一个人,访员有一个事先准备好的表格,他要把家庭中居住的所有人都系统地填在上面,通常要按照其年龄和性别排序,每个人都被赋予一个序列号。选中谁进行访谈取决于家庭中的总人数。Kish 表包括了一个选择表,它会告诉访员应选择哪一个编号的个人。访员只需要从表中列有家庭中人数的那一列中读出他应该进行访问的那个序列号就可以了。Kish 表是被广泛应用的标准技术,可以适用于不同调查设计的需要。

——Heather Laurie

(高勇译校)

参考文献

Kish, L. (1949). A procedure for objective respondent selection within a household. *Journal of the American Sociological Association*, *44*, 380-387.

Kish, L. (1965). *Survey sampling*. New York: John Wiley.

Kolmogorov-Smirnov 检验(Kolmogorov-Smirnov Test)

有两种 Kolmogorov-Smirnov 检验,一种是单样本拟合优度【Goodness-of-Fit】检验,一种是双样本检验。

单样本检验

用于定序数据的单样本检验有许多种

(Gibbons & Chakraborti, 1992; Siegel & Castellan, 1988),在不同应用情境中各有优势。Kolmogorov-Smirnov 单样本检验被认为比卡方拟合优度检验具有更强效力,尤其是样本量较小时。这种检验适用于定序数据,虽然它实质上检验的是样本数据分布和总体数据分布的匹配程度;因此,总体分布可包括正态分布之外的其他分布形式。它也假定潜在的是连续分布,虽然数据可能呈现出区间不等的分段(参见 Gibbons & Chakraborti, 1992),这导致了其他算法,如下所示:

检验本身比较的是数据的累积分布和期望总体分布的累积分布。它假定存在一个潜在的连续分布,样本量较小时肯定优于卡方检验【Chi-Square Test】,因为卡方检验受限于最小期望值。两个分布的最大差异幅度就是检验的因素,它就是两个累积的相对频率(即在整体中的占比)之间的最大绝对差值。然后再用不同 α 值下的最大差异表进行比照检验。

正确应用这些检验方法的难点在于序列区间的性质。文献中讨论了这一问题的两种基本类型,每种情况下 Kolmogorov-Smirnov 检验的用法有所不同。在某种程度上,每种问题类型都是检验的不同应用方式:

(1)序列区间并没有事先定义好,潜在的分布是假定的;因此,检验是一个连续函数(如正态分布或均匀分布)用数理方法决定的(参见 Gibbons & Chakraborti, 1992)。样本数据是分段的,而总体“数据”是连续的。

(2)区间是事先定义好且互斥的,也存在分区间的总体数据;因此,总体数据是分段的,样本数据也是分段的(参见 Siegel & Castellan, 1988)。这使得检验等同于两个样本都是分段时的双样本检验。

下面用一些例子来说明其具体用法。

从数据分布中得到连续总体数据

例如,问题如下:某个学校的工资分布是否和整个县的工资分布相同呢?在此假定全县的工资服从正态分布。表 1 中显示了这个学校的工资频数,样本数据的相应累积相对频数,此外,还提供了基于正态分布的总体累积相对分布。图 1 用图形方式显示了上述信息。显然对每个数据点来说,有两个可以进行比较的点:上区间和下区间。在图中,有两个一端在连续总体分布上的箭头,总体分布可用来和样本数据中的两个分段之一进行匹配。

因此,Kolmogorov-Smirnov 检验需要计算所有点上的这两个差异,结果显示在表 1 的最后两列中。然后比较从全部差异中得到的最大差异和相应 α 值下的临界差异。表 2 显示了近似估计的临界值 $D_{临界值}$。

在本例中,如果 α 值被选为 0.05,那么允许的最大差异值为:

$$D_{临界值} = \frac{1.36}{\sqrt{20+1}} = 0.297$$

表 1　Kolmogorov-Smirnov 单样本检验数据,数据是有序列的,总体分布是数学函数

学校工资	f_o	CF_o 学校	CF_e 全县	$CF_{oi}-CF_{ei}$ 临近行	$CF_{o(i-1)}-CF_{ei}$ 偏移行
9 278	2	0.100	0.131	−0.031	−0.131
10 230	1	0.150	0.168	−0.018	−0.068
11 321	1	0.200	0.218	−0.018	−0.068

续表

学校工资	f_o	CF_o 学校	CF_e 全县	$CF_{oi}-CF_{ei}$ 临近行	$CF_{o(i-1)}-CF_{ei}$ 偏移行
13 300	1	0.250	0.326	−0.076	−0.126
14 825	2	0.350	0.422	−0.072	−0.172
16 532	2	0.450	0.535	−0.085	−0.185
18 545	1	0.500	0.664	−0.164	−0.214
21 139	3	0.650	0.804	−0.154	−0.304
22 145	2	0.750	0.847	−0.097	−0.197
24 333	2	0.850	0.918	−0.068	−0.168
28 775	1	0.900	0.983	−0.083	−0.133
30 987	1	0.950	0.994	−0.044	−0.094
32 444	1	1.000	0.997	0.003	−0.047
合计	20				

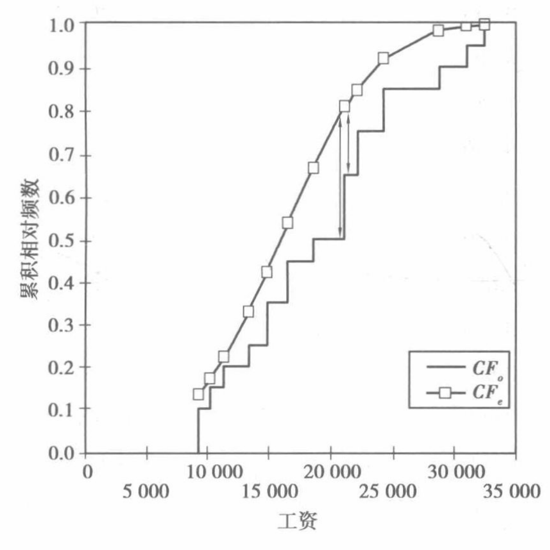


图 1　观察和期望累积相对频数分布

注:这些频数代表着总体分布是假定的,是从数学函数(在此即正态分布)中得到的情形。

现在再来看表 1,从最后两行可以看出,最大差值 0.305 已超过了临界值,学校的分布和假想的全县分布之间是有着显著差异的。布莱克(Black,1999)的书中对同一数据比较了这一检验和卡方检验、t 检验【t-Test】。需要注意的是,对某一个数据进行这三种检验,不仅其效力(power)水平不同,而且要回答的问题也是不同的。

表 2　用于 Kolmogorov-Smirnov 检验的临界最大差异 $D_{临界值}$

	α		
样本量	0.10	0.05	0.01
$25>n>10$	$\frac{1.22}{\sqrt{n+1}}$	$\frac{1.36}{\sqrt{n+1}}$	$\frac{1.63}{\sqrt{n+1}}$
$n\geqslant 25$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

从完整调查中得到分段的总体数据

在这种情况下,样本数据是和真正的总体数据相比,两者中每个定序类别的频数都是已知的。为了说明这一点,我们使用前一个例子中的样本数据,但是把它和全县教师的工资记录相比。图 2 显示了数据和相应的分段图,很明显,只有区间内的数据可进行比较。例如,双箭头表明了一对区间之间的差异。因此,Kolmogorov-Smirnov 单样本检验只需要使用累积相对频数中的临近行差值(表 3),最大差值将用来和选定的 α 值下的最小临界值进行比较。最大差值为 0.278,它小于前面发现的 0.297(正负号在此没有关系),因此,不存在显著差异。

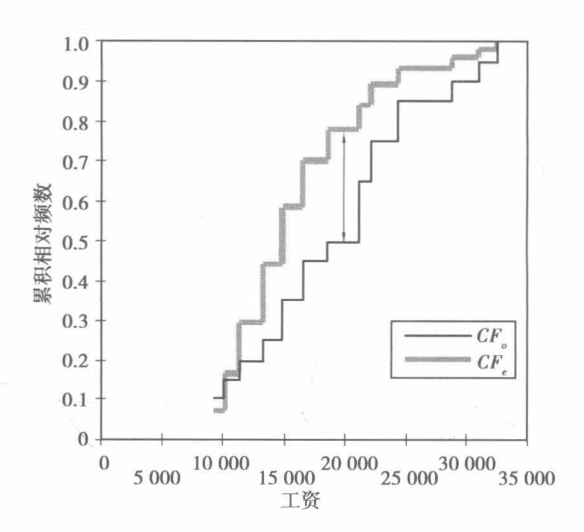


图 2　观察和期望累积相对频数分布

注:这些分布对应于总体数据和样本数据都是分段时的情形(即总体数据来自直接调查)。

表 3　Kolmogorov-Smirnov 单样本检验数据,数据是有序列的,总体分布来自定序的样本数据

学校工资	f_o 学校	f_o 全县	CF_o 学校	CF_e 全县	$CF_{oi}-CF_{ei}$ 临近行
9 278	2	260	0.100	0.097	0.003
10 230	1	287	0.150	0.205	−0.055
11 321	1	320	0.200	0.325	−0.125
13 300	1	350	0.250	0.456	−0.206
14 825	2	365	0.350	0.593	−0.243
16 532	2	287	0.450	0.700	−0.250
18 545	1	208	0.500	0.778	−0.278
21 139	3	145	0.650	0.832	−0.182
22 145	2	125	0.750	0.879	−0.129
24 333	2	100	0.850	0.916	−0.066
28 775	1	85	0.900	0.948	−0.048
30 987	1	74	0.950	0.976	−0.026
32 444	1	64	1.000	1.000	0.000
合计	20	2 670			

如果数据并不服从事先选定的分布,这种任意分布的(distribution-free)比较在很多情况下就有很大优势。在 SPSS 中要进行等价检验,必须使用双样本功能,其中一个的

数据值是总体数据。

对两个独立样本进行检验

涉及两组间进行比较的研究设计并不总是生成区间数据或比值数据，然后用 t 检验的方式来检验假设。虚无假设 H_0 可以是两组间因变量在不同类别或水平上的频数分布性质是相同的。换言之，两个组的不同特征（自变量【Independent Variable】）是否影响到了因变量【Dependent Variable】的不同类别或水平上的分布性质呢？

Kolmogorov-Smirnov 检验会查看这两个样本是否是从同一个总体中抽取的，这需要假定存在一个潜在的连续分布。检验程序与前面所述基本相同，只是这次比较的是临近行的累积相对频数。当 $n>25$ 且 $m>25$ 时，可以用以下方式计算出允许的最大差值：

$$D(0.05) = 1.36\sqrt{\frac{m+n}{mn}} \qquad (1)$$

对于小样本来说，需要用到专用的临界值表（参见 Siegel & Castellan, 1988），或者用 SPSS 可以自动给出概值。西格尔和卡斯特兰（Siegel & Castellan, 1988）认为这种检验的效力比 Wilcoxon 检验【Wilcoxon Test】或 Mann-Whitney U 检验【Mann-Whitney U Test】略微强一些，但是如果样本量较大，结论则会相反。

——Thomas R. Black

（高勇译校）

参考文献

Black, T. R. (1999). *Doing quantitative research in the social sciences: An integrated approach to research design, measurement, and statistics.* London: Sage.

Gibbons, J. D., & Chakraborti, S. (1992). *Nonparametric statistical inference* (3rd ed.). New York: Marcel Dekker.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Duxbury.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.

Kruskal-Wallis H 检验（Kruskal-Wallis H Test）

这一检验也被称为对序列数据（ranked data）进行的 Kruskal-Wallis 单因素方差分析【One-Way, ANOVA】，它根据中位数【Median】来确定三个或多个独立样本是否属于同一总体【Population】。它可以类比于参数检验中的单因素方差分析，是它在双样本情境中的对等检验。其虚无假设【Null Hypothesis】是中位值都与总体中位值相同，统计检验用来检查考虑到样本中的自然波动之后，样本中位值是否都足够相近。数据应该至少是定序的，假定存在一个潜在的连续分布（Siegel & Castellan, 1988）。

和其他基于序列的检验一样，这种检验也要求所有样本的所有原始数据要进行排序，然后赋予一个秩（rank），最小的赋秩为 1，一直到 N（所有 k 个样本中的所有独立观察数）。每个样本的所有秩进行加总，然后计算每组的平均秩。这里的假定是，如果这些样本是从同一个总体中抽取出来的，那么这些平均秩在考虑自然波动之后也应非常接近。在如下 Kruskal-Wallis 统计量的表达式中可以很容易地看出这点，这一表达式会让人联想起方差的计算：

$$H = \frac{12\sum_{j=1}^{k} n_j(\bar{R}_j - \bar{R})^2}{N(N+1)}$$

这一统计量也可用以下表达式计算，这个表达式更容易从原始数据中算出（Howell，2002），特别是在电子表格当中（Black，1999）。

$$H = \left[\frac{12}{N(N+1)}\sum_{j=1}^{k}\frac{R_j^2}{n_j}\right] - 3(N+1)$$

式中 R_j——第 j 组的秩和，$j=1,\cdots,k$；

$\bar{R}_j$——第 j 组的平均秩；

$\bar{R}$——所有秩的总均值；

n_j——第 j 组的观察数；

N——所有组的总观察数。

如果组数大于3，且每组中的观察数大于4时（或者组数为3但样本量较大时），抽样分布就接近于卡方分布【Chi-Square Distribution】。如果存在大量的同分（tie）时（大于样本量的25%），H 统计量的值就会增大，因此，需要进行一个修正（参见 Siegel & Castellan，1988）。如果 R_j 的所有值都彼此相等，那么 H 统计量为0。

例如，表1中的数据显示了人们对一种名叫“克瑞皮斯（Creepies）”的新型早餐麦片的打分，这种麦片是巧克力口味的，形态像各种昆虫。每个人会用四个5分量表对麦片进行打分：质地、口味、颜色、形状。因此，可能的得分范围是4~20。有人猜测成人可能不像儿童一样喜欢它。样本被抽取出来，以避免儿童和成人来自相同家庭。首先，对打分进行了排序；然后，计算 H 统计量：

$$H = \left[\frac{12}{24\times(24+1)}\sum_{j=1}^{3}\frac{R_j^2}{n_j}\right] - 3\times(24+1)$$

$$= \left[\frac{12}{600}\times 4\ 324.2\right] - 75$$

$$= 11.48$$

与临界值 $\chi^2(df=2)=5.99$ 相比，三个组是存在差异的。

表1 三个组对“克瑞皮斯”早餐麦片的打分

男性		女性		儿童	
打分	秩	打分	秩	打分	秩
12	8.5	13	6.5	20	1
10	12.5	12	8.5	19	2
7	16.5	11	10.5	18	3.5
7	16.5	10	12.5	18	3.5
6	18.5	9	14.5	16	5
5	20.5	6	18.5	13	6.5
4	23	5	20.5	11	10.5
4	23	4	23	9	14.5
R_1	139		114.5		46.5
n_1	8		8		8
R_i^2/n_i	2 415.13		1 638.78		270.28
R_i/n_1	17.38		14.31		5.81
			男性-儿童	女性-儿童	男性-女性
$\alpha/k(k-1)$	0.008 3	$\Delta(R/n)$	11.56	8.50	3.06
$z(0.008\ 3)$	2.394	$\Delta(R/n)$临界值	8.46	8.46	8.46

注：平分时，秩就是相邻秩的平均值。$N=24$；$H=11.48$；$\alpha=0.05$；$\chi^2(2)=5.99$。

对 Kruskal-Wallis H 检验的事后分析

进行成对比较就是对各对平均秩之间的差异进行检验,标准需要进行调整以便考虑到它们之间并不独立。如果样本量较大,它们的差异是近似服从正态分布的(Siegel & Castellan, 1988),其临界值【Critical Value】为:

$$\Delta\left(\frac{R_i}{n_i}\right)_{\text{临界值}} = z \times \frac{\alpha}{k(k-1)}\sqrt{\frac{N(N+1)}{12}\left(\frac{1}{n_a}+\frac{1}{n_b}\right)}$$

如果所有组的样本量都相同,那么就需要进行一次这样的计算。如果比较只是在每个处理组和一个对照组之间进行,而不在处理组之间进行,那么还需进行调整(类似于 Dunnett 对 ANOVA 的事后检验)(参见 Siegel & Castellan, 1988)。对于前面的例子来说:

$$\Delta\left(\frac{R_i}{n_i}\right)_{\text{临界值}} = z \times \frac{0.05}{3\times(3-1)}\sqrt{\frac{24\times(24+1)}{12}\times\left(\frac{1}{8}+\frac{1}{8}\right)}$$

$$= 2.394\sqrt{12.5} = 8.46$$

从表 1 中可以看出,成年男性与儿童之间的差异和成年女性与儿童之间的差异都大于 8.46,因此差异都是显著的。在男性与女性之间不存在显著差异。因此,这证实了前面的推测。

——Thomas R. Black

(高勇译校)

参考文献

Black, T. R. (1999). *Doing quantitative research in the social sciences: An integrated approach to research design, measurement, and statistics*. London: Sage.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Duxbury.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.

峰态(Kurtosis)

峰态系数、偏态系数及其标准差是用来测量分布形状的常用单变量描述性统计量。峰态系数是指一个分布的峰度(peakedness),测量的是分布的尾部和它最众多的取值之间的关系。峰态系数常常用来筛选数据是否服从正态分布。偏态值测量的是分布的对称性,峰态值测量的是顶峰性,它们通常一起进行解释。如果数据服从正态分布,根据统计学惯例,偏态系数和峰态系数都应当在 -2 ~ +2 中。其数据公式为:

$$\text{峰态系数} = \frac{\sum (X-\mu)^4}{N\sigma^4} - 3$$

式中 X——分布中的个体得分;

μ——样本均值;

N——数据点的数目;

σ——方差的平方根。

从数学上来讲,峰态系数是四阶【Order】中心矩【Moment】。标准正态分布的峰态系数为 3。公式中从峰态系数中减去 3,这样峰态系数为 0 就成为了标准正态

分布的标志。

在分析峰态时一定要确保足够的样本，因为对于小样本来说，峰态系数的大小和方向都会不稳定。如果发现数据不服从正态分布，研究者可能会希望用数据转换【Transformation】的方式来产生一个正态分布。统计程序通常会输出直方图、箱形图【Boxplot】、箱须图、茎图、茎叶图【Stem-and-Leaf】来展示分布的方向，以及偏态系数和峰态系数的大小。如果数据不服从正态分布，就可以用这些图来确定使用何种转换技术。一个简单的例子就是，可以在直方图中加入正态曲线来评估分布的形状。偏态系数和峰态系数为正时，对数转换和平方根转换是最常见的修正方法；偏态系数和峰态系数为负时，平方根转换是常见的修正方法。立方根转换和负反根转换（negative reciprocal root transformation）分别用于修正极端的负峰态系数和正峰态系数（Newton & Rudestam, 1999）。

数据根据不同的分布形状可以是尖峰的、平峰的、常峰的（Newton & Rudestam, 1999）。峰态系数为大的正值表示尖峰，分布有一个非常尖的峰，数据聚集在均值周围，尾部更厚或更长，这意味着尾部有很少的案例。如图 1 中的假想数据所示，与正态分布数据相比，尖峰数据聚集在均值附近，尾部有很少的案例，结果就是中等的正峰态系数。

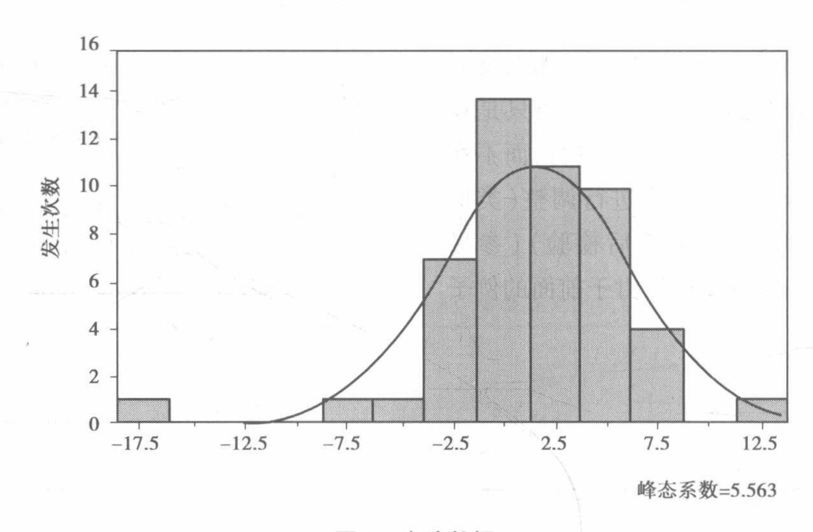


图 1　尖峰数据

偏态系数和峰态系数为正表示均值比中位值大。峰态系数为大的负值表明数据呈现平峰，直方图会表明分布很平，数据并不聚集在均值周围，有着更短或更轻的尾部，这意味着尾部比正态分布中要有更多的案例。偏态系数和峰态系数为负表明中位值比均值大。常峰数据近似于正态分布，其峰态系数可以是一个较小的正值或较小的负值。偏态系数和峰态系数较小表明均值和中位值相对接近。

——Amna C. Cameron

（高勇译校）

参考文献

Newton, R. R., & Rudestam, K. E. (1999). *Your statistical consultant: Answers to your data analysis questions*. Thousand Oaks, CA: Sage.

L

实验室实验(Laboratory Experiment)

所有的实验【Experiment】都涉及把参与者随机分配到研究条件中,以及自变量【Independent Variable, IV】和因变量【Dependent Variable, DV】。实验既可以在高度受控的实验室条件中进行,也可以在控制较少的实地情境中进行。两种情况各有利弊。在实验室中,研究者能够控制可能影响参与者的所有外在因素——对条件有着严格控制。实地实验则并非如此,因此在这些情境中结论有时更不可靠,因为可能有非控制的变异在施加影响。

实验过程中的随机分配意味着从事先定义好的备选集合中,所有的参与者有同等机会被选中进入某一种或另一种研究条件中。在随机分配之后,被选中进入不同条件的参与者接受不同水平的IV。在某些设计中,参与者将接受前试(即用理论上来说对IV不同水平引发的变异敏感的量度来进行评估)。在其他设计中,只在处理后测量DV(参见Crano & Brewer, 2002)。在施加处理之后,再次对处理组和控制组(不接受实验处理的组)进行评估,如图1所示。

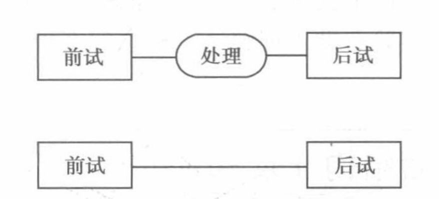


图1 前试/后试/控制组实验设计的例子

注:如果处理组和控制组不进行前测,那就是只有后试/控制组的实验设计。每种实验设计中,参与者均被随机分配到不同条件中。

在进行实验时,作为研究核心的理论构念,以及这种概念化【Conceptualization】的测量,必须用可观察的行为或事件来定义。这个过程被称为操作化(operationalization)。然后,参与者被随机分配到用IV不同水平定义的实验条件中。有时,使用不止一个IV;在此类情形中,通常一个IV的每个水平都会与另一个IV的每个水平进行综合(combine),生成析因设计【Factorial Design】。在理想状态下,同等数量的参与者被分配到因素综合生成的每种条件中。在分配之后,参与者被施加以特定的处理(IV),或者处理的交互综合中(如果用到多个IV),然后观察这种施加对DV的影响。

实验室实验是高度受控的研究技术。通过随机化,理论上来讲参与者在各个方面都是等同的,除了有的被施加以IV,有的则没有(或者它们被系统地施加以IV的不同值或不同水平)。因此,除非IV有影响,它们不会在DV上显示出差异。实验室实验研究的高度受控特征也使得有人质疑这种情景中得到的结论的概推性。反过来讲,实

验室实验的随机化和受控性也使得研究能够排除那些虽然影响 DV 但却没有考虑外在变量的影响。实验室实验最适合用何种情况，这要取决于概推性与受控性之间的权衡。

实验室实验是确定因果关系的理想方法。如果对要研究的现象已经有相当的了解，使用这种方法的效果最好。对于提出假设来说，它不是一种特别有效的方法。但是如果围绕某一现象的理论已经相对成熟，在理论基础上已发展出了精细的假设，那么实验室实验有助于理解这一构念的准确本质。如果对具体可预测的因果关系存在不同解释，它是检验这些解释的理想方法，可以确定哪些因素对给定结果有放大作用或阻碍作用。

——William D. Crano

（高勇译校）

参考文献

Crano, W. D., & Brewer, M. B. (2002). *Principles and methods of social research* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

滞后结构（Lag Structure）

时间序列【Time-Series】模型通常假定自变量对因变量既有即时影响，也有延时影响。这种关系可以用滞后结构来加以建模，其参数可以用一些合适的方法来加以估计。如果只有一个自变量 x，而且忽略误差项，最一般的滞后结构为：

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \cdots$$

估计上述模型的参数是不可能的，因为存在无限个参数。

有限分布滞后结构假定在给定的滞后期（如 K）之后，x 就不再影响 y 了。这时，滞后结构就只有有限个参数，即 $\beta_{K+1} = \beta_{K+2} = \cdots = 0$。虽然在原则上这是可以进行估计的，但多重共线性【Multicollinearity】使得每个参数的估计都很困难，除非假定 x 对 y 的影响只有 1 或 2 个滞后期（即使如此多重共线性仍然可能是个问题）。因此，分析者通常会对 β 施加一些参数结构。最常见的假定是它们遵循多项式结构，如随时间呈现二项式结构，即 $\beta_k = \lambda_0 + \lambda_1 k + \lambda_2 k^2$。这一约束意味着分析者只需估计三个参数。

许多研究者更愿意考虑 x 对 y 的影响有无限个滞后期（或者至少是对滞后期没有限定），他们认为不应该假定效应只延续到事先确定的滞后期，然后效应就变成零了。这种无限分布滞后的最常见参数化方法是几何滞后结构（geometric lag structure），即 x 对 y 的影响被假定呈现几何衰减，即 $\beta_k = \beta\rho^k$。这可以用代数方法转化为一个有滞后因变量的模型。认为 x 对 y 的影响随滞后长度呈现几何衰减的假定通常是合理的。如果这一约束条件太强，研究者可以结合有限和无限滞后结构，即最初几个滞后期如有限滞后结构模型中那样进行自由估计，而随后的滞后期被假定呈现几何衰减。这样就在有限滞后模型的灵活性和无限几何滞后模型的简洁性之间实现了良好的平衡，不会强迫分析者估计太多的参数，也能够避免多重共线性问题的麻烦。

如果有多个自变量，研究者必须在让它

们全都遵循同一滞后结构的简洁性和允许每个变量有着独特滞后结构的灵活性之间进行选择。选择灵活性的代价是分析者必须估计许多参数。分析者也可使用滞后结构来对误差过程进行建模。时间序列分析者可以利用的各种工具,如相关图(correlograms),应当与理论洞察结合起来,以使分析者选择出最佳的滞后结构。

——Nathaniel Beck

(高勇译校)

参考文献

Beck, N. (1991). Comparing dynamic specifications: The case of presidential approval. *Political Analysis*, *3*, 51-87.

Harvey, A. (1990). *The econometric analysis of time series* (2nd ed.). Cambridge: MIT Press.

Mills, T. C. (1990). *Time series techniques for economists*. Cambridge, UK: Cambridge University Press.

Lambda(λ)

λ 系数是对定类【Nominal】变量之间关系强度的测量。λ 系数也被称为 Γ 预测系数,它测量的是如果知道了定类预测变量【Predictor Variable】的类别,能够在多大程度上提升对因变量【Dependent Variable】的预测。它是一个误差消减比例测量,因为它显示了知道一个变量的取值可以多大程度上减少对另一个变量的预测误差。λ 系数是非对称性系数,因为它的取值取决于两个变量中何者被作为因变量,如果一个变量是因变量,另一个变量是自变量,就会得到不同的 λ 系数值。自变量和因变量互换会改变 λ 系数的取值:

$$\lambda = \frac{\sum f_i - M}{N - M}$$

其中的 f_i 是自变量的第 i 个类别中最大的单元格频数,M 是因变量的最大边缘频数,N 是总案例数。分母是不知道自变量类别时预测因变量所犯的错误数,分子是知道自变量类别时预测因变量能够少犯的错误数。

λ 系数的取值范围为 0~1,不会取到负值。多数关系测量将统计独立【Independence】视为两个变量间关系为零的定义,但 λ 系数在自变量每个类别中因变量的众数回答都相同时变量间关系为零。例如,分析一个人的居住地区和他最喜欢的颜色的关系时,如果在每个地区多数人都选择蓝色作为自己最喜欢的颜色,那么 λ 系数的取值为 0。

如果两个变量是统计独立的,那么 λ 系数的取值也为 0。

表 1 最喜欢的颜色和居住地区的关系——无助于预测

	东北	南方	西方
蓝色	60	40	80
红色	20	30	10
灰色	20	30	10

注:λ = 0(用居住地区预测最喜欢的颜色)。

表 2　最喜欢的颜色和居住地区的关系——完全预测

	东北	南方	西方
蓝色	100	0	0
红色	0	0	100
灰色	0	100	0

注：$\lambda=1$（用居住地区预测最喜欢的颜色）。

λ 系数在“条件性全体一致（conditional unanimity）”的情况下取值为 1。也就是说，它将两个变量间存在的完美关系解释为：在自变量的每个类别中，所有案例都落入因变量的同一类别中。如果东北地区的人都选择蓝色作为最喜欢的颜色，南方人都选择灰色，西部人都选择红色，那么知道了一个人的居住地区就可以完全预测其喜欢的颜色，因此，λ 系数等于 1。

也有对称性的 λ 系数，不论哪个变量作为因变量，这种系数都能得到相同的取值。

——Herbert F. Weisberg
（高勇译校）

参考文献

Blalock, H. M. (1979). *Social statistics* (Rev. 2nd ed.). New York: McGraw-Hill.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classification. *Journal of the American Statistical Association*, *49*, 732-764.

Guttman, L. (1941). An outline for the statistical theory of prediction. In P. Horst (with collaboration of P. Wallin & L. Guttman) (Ed.), *The prediction of personal adjustment* (Bulletin 48, Supplementary Study B-1, pp. 253-318). New York: Social Science Research Council.

Weisberg, H. F. (1974). Models of statistical relationship. *American Political Science Review*, *68*, 1638-1655.

潜约束分析（Latent Budget Analysis）

一个约束（budget），表示为 $\mathbf{p}$，是指一个由非负元素构成的向量，这些元素加总等于一个固定常数 c。一个约束的例子是 J 个互斥且穷尽的活动（如睡眠、工作、家务等）在一天中的时间分配。另一个约束的例子是一笔 20 000 元的拨款在 J 个文化项目上的分配。注意，时间约束的各个部分加总起来等于 24 小时或 1 440 分钟，财政约束的各个部分加总起来等于 20 000 元。不失一般性地，我们可以将一个约束的各个部分除以固定常数 c，以使它们加总等于 1。I 个（或 I 组）被访者的时间约束（即 $\mathbf{p}_1,\cdots,\mathbf{p}_I$）或者 I 个 20 000 元拨款的财政约束可以表示为 $I\times J$ 的数据矩阵。这类数据被称为构成数据（compositional data）。潜约束分析的理念在于，将 I 个约束写为更少数目的潜约束（latent budgets）的混合，这些潜约束用 $\beta_1,\cdots,\beta_k$ 表示，其中的 $k<I$。

令 Y_E 为解释变量（如国籍、社会经济地位、时间点）或者有 I 个类别的指标变量，用

i 标记。令 Y_R 为有着 J 个类别的响应变量，用 j 标记，令 X 为有 K 个类别的潜变量，用 k 标记。约束可以写作以下形式：

$$\mathbf{p}_i = [P(Y_R = 1 \mid Y_E = i), \cdots, P(Y_R = j \mid Y_E = i), \cdots, P(Y_R = J \mid Y_E = i)]$$

其中的 $i = 1, \cdots, I$。潜约束可以写作：

$$\beta_k = [P(Y_R = 1 \mid X = k), \cdots, P(Y_R = j \mid X = k), \cdots, P(Y_R = J \mid X = k)]$$

其中的 $k = 1, \cdots, k$。令 $P(X = k \mid Y_E = i)$ 表示一个权重，这个权重表明了潜约束 β_k 决定的预算比例 $\mathbf{p}_i$。在潜约束分析中，约束 $\mathbf{p}_i(i = 1, \cdots, i)$ 通过期望约束（expected budgets）π_i 进行估计，以使：

$$\mathbf{p}_i \approx \pi_i = \sum_{k=1}^{k} P(X = k \mid Y_E = i)\beta_k$$

构成数据可以用潜约束 $\beta_i, \cdots, \beta_k$ 以如下程序来进行解释：首先，潜约束 $\beta_1, \cdots, \beta_k$ 被解释和命名。比较约束的构成部分和独立约束的构成部分 $[P(Y_R = 1), \cdots, P(Y_R = j)]$ 有助于提取出潜约束的特征来。例如，如果 $P(Y_R = j \mid X = k) > P(Y_R = j)$，那么 β_k 的特征就是比平均情况下更多倾向于组成部分 j。其次，在已命名的潜约束基础上来解释约束 $\pi_1, \cdots, \pi_I$，方法是比较权重 $P(X = K \mid Y_E = i)$ 与平均权重

$$P(X = k) = \sum_{i=1}^{I} P(X = k \mid Y_E = i)P(Y_E = i)$$

例如，如果 $P(X = k \mid Y_E = i) > P(X = k)$，那么约束 π_i 的特征就是比平均情况下更多地接近于潜约束 β_k。

潜约束分析的理念是由哥德曼（Goodman, 1974）年提出的，他称之为非对称性的潜类分析【Latent Class Analysis】；这一理念也由德·莱乌和范·德·希伊登（de Leeuw & van der Heijden, 1988）独立提出，他们称之为潜约束分析。与潜约束分析相关的议题在下列论文中进行了讨论：de Leeuw & van der Heijden（1998，图形表示）、van der Ark, van der Heijden, & Sikkel（1999，模型识别）、van der Heijden, Mooijaart, & de Leeuw（1992，约束潜约束模型）、van der Heijden, van der Ark, & Mooijaart（2002，应用）。

——L. Andries van der Ark

（高勇译校）

参考文献

de Leeuw, J., & van der Heijden, P. G. M. (1988). The analysis of time-budgets with a latent time-budget model. In E. Diday (Ed.), *Data analysis and informatics* (Vol. 5, pp. 159-166). Amsterdam: North-Holland.

Goodman, L. A. (1974). The analysis of systems of qualitative variables when some of the variables are unobservable: I. A modified latent structure approach. *American Journal of Sociology*, *79*, 1179-1259.

van der Ark, L. A., & van der Heijden, P. G. M. (1998). Graphical display of latent budget analysis and latent class analysis, with special reference to correspondence analysis. In M. Greenacre & J. Blasius (Eds.), *Visualization of categorical data* (pp. 489-508). San Diego: Academic Press.

van der Ark, L. A., van der Heijden, P. G. M., & Sikkel, D. (1999). On the identifiability of the latent model. *Journal of Classification*, *16*, 117-137.

van der Heijden, P. G. M., Mooijaart, A., & de Leeuw, J. (1992). Constrained latent budget analysis. In P. Marsden (Ed.), *Sociological methodology* (pp. 279-320). Cambridge, UK: Basil Blackwell.

van der Heijden, P. G. M., van der Ark, L. A., & Mooijaart, A. (2002). Some examples of latent

budget analysis and its extensions. In J. A. Hagenaars & A. McCutcheon (Eds.), *Applied latent class analysis* (pp. 107-136). Cambridge, UK: Cambridge University Press.

潜类分析(Latent Class Analysis)

潜类(LC)分析的基本理念很简单:存在着一些不可观察的小组,一个认定的统计模型的某些参数在这些分组中有所不同。这些小组构成了一个类别潜变量的不同类别(参见潜变量【Latent Variable】)。这个基本理念有几个看似无关的应用,最重要的包括聚类、量表、密度估计、随机效应模型。在社会科学之外,LC 模型常被称为有限混合模型(finite mixture models)。

1950 年代,拉扎斯菲尔德(Lazarsfeld)引入了 LC 分析,他用这种技术基于一些二分的可观察变量构建类型(或聚类)。20 多年后,哥德曼(Goodman)发明了一种算法来获取模型参数的最大似然估计,从而使模型能够在实践中应用。他还将模型扩展到了多项显变量和多个潜变量的情形中,在模型识别问题上也作出了重要贡献。与此同时,哈伯曼(Haberman, 1979)说明了 LC 模型与有缺失单元格频数的列联表的对数线性模型之间存在关联。此后又提出了经典 LC 模型的许多重要扩展,如包括(连续)协变量的模型、局部相依、定序变量、多个潜变量、重复测量等。哈格纳尔斯(Hagenaars, 1990)为存在离散潜变量的类别数据分析提出了一个一般性框架,费蒙特(Vermunt, 1997)对此进行了扩展。

在社会科学中,LC 分析和有限混合模型主要被认为是分析类别数据的工具,但是它们在其他几个领域也同样有用。例如,在密度估计中,可以利用复杂密度,也可以用简单密度的有限混合来近似这一事实。LC 分析也可以用于对连续的观察变量进行概率性聚类分析,这种方法比传统的聚类技术(如 K-means聚类)更为优越(参见潜相模型【Latent Profile Model】)。另一应用领域是处理不可观察的异质性,例如对彼此相依的观察进行回归分析(参见非参数随机效应模型【Nonparametric Random-Effect Model】)。

分析类别指标的经典潜类模型

令 X 代表潜变量, Y_l 是 L 个观察变量或显变量之一,其中 $1 \leqslant l \leqslant L$。此外,令 C 为潜变量数目, D_l 是 Y_l 的类别数目。一个特定的潜类可以用 x 列举,$x = 1, 2, \cdots, C$, Y_l 的特定值可以用 y_l 来列举, $y_l = 1, 2, \cdots, D_l$。向量符号 $\mathbf{Y}$ 和 $\mathbf{y}$ 用于指涉完整的响应模式(response pattern)。

为了具体说明,我们可以考虑从 1987 年综合社会调查(General Social Survey)中得到的如表 1 所示的小数据集:

表 1

Y_1	Y_2	Y_3	频 数	$P(X=1\|\mathbf{Y}=\mathbf{y})$	$P(X=2\|\mathbf{Y}=\mathbf{y})$
1	1	1	696	0.998	0.002
1	1	2	68	0.929	0.071
1	2	1	275	0.876	0.124

续表

Y_1	Y_2	Y_3	频　数	$P(X=1\mid\mathbf{Y}=\mathbf{y})$	$P(X=2\mid\mathbf{Y}=\mathbf{y})$
1	2	2	130	0.168	0.832
2	1	1	34	0.848	0.152
2	1	2	19	0.138	0.862
2	2	1	125	0.080	0.920
2	2	2	366	0.002	0.998

三个二分指标 Y_1, Y_2, Y_3 是对以下三句话的回答:"是否允许反宗教者发表言论"(1=允许,2=不允许);"是否允许反宗教者教书"(1=允许,2=不允许);"是否应从图书馆中撤出反宗教书籍"(1=不应撤出,2=应撤出)。利用 LC 分析,可以识别出对反宗教者的容忍度有所不同的群体。

任何类型的 LC 模型的基本理念就是,响应模式 $\mathbf{y}$ 的概率 $P(\mathbf{Y}=\mathbf{y})$ 是 C 个分类别的概率 $P(\mathbf{Y}=\mathbf{y}\mid X=x)$ 的加权平均,即:

$$P(\mathbf{Y}=\mathbf{y}) = \sum_{x=1}^{C} P(X=x)P(\mathbf{Y}=\mathbf{y}\mid X=x) \tag{1}$$

在此,$P(X=x)$表示所有人中属于 LC x 的比例。

在经典 LC 模型中,这个基本理念是与局部独立【Local Independence】假定结合在一起的。L 个显变量被假定在每个 LC 当中是彼此独立的,这一点可以表达如下:

$$P(\mathbf{Y}=\mathbf{y}\mid X=x) = \prod_{l=1}^{L} P(Y_l=y_l\mid X=x) \tag{2}$$

在估计出条件响应概率 $P(Y_l=y_l\mid X=x)$ 后,比较不同潜类的这些概率可以显示出潜类之间的差异,可用于给这些潜类命名。将两个基本公式(1)和(2)组合,可以得到以下对 $P(\mathbf{Y}=\mathbf{y})$的模型:

$$P(\mathbf{Y}=\mathbf{y}) = \sum_{x=1}^{C} P(X=x)\prod_{l=1}^{L} P(Y_l=y_l\mid X=x)$$

用上述例子中的小数据估计一个两类模型,可得到如表 2 所示结果:

表 2

	$X=1$(宽容)	$X=2$(不宽容)
$P(X=x)$	0.62	0.38
$P(Y_1=1\mid X=x)$	0.96	0.23
$P(Y_2=1\mid X=x)$	0.74	0.04
$P(Y_3=1\mid X=x)$	0.92	0.24

这两个潜类分别包括了62%和38%的个体。第一类别可以称为"宽容"类,因为与属于第二类别"不宽容"的人相比,属于第一类别的人在所有指标上有更高概率选择更为宽容的回答。

与聚类分析【Cluster Analysis】相类似,LC分析的目标之一是将个体分配到每个潜类上。属于某个LC x 的概率——常被称为后验归属概率(posterior membership probability)——可用贝叶斯法则得到:

$$P(X=x \mid \mathbf{Y}=\mathbf{y}) = \frac{P(X=x)P(\mathbf{Y}=\mathbf{y} \mid X=x)}{P(\mathbf{Y}=\mathbf{y})} \quad (3)$$

最常见的分类准则是众数分配,相当于把个体分配到 $P(X=x \mid \mathbf{Y}=\mathbf{y})$ 最高的那个LC中。第一个表中显示的类别归属概率表明,那些至少有两个宽容性回答的人被归类到了"宽容"类别中。

潜类模型的对数线性表达式

哈伯曼(Haberman, 1979)表明,LC模型可以设定为一个有缺失单元格的列联表的对数线性模型【Log-Linear Model】,或者更为准确地说,一个包括潜变量 X 作为额外维度的扩展表的对数线性模型。对于 $P(X=x \mid \mathbf{Y}=\mathbf{y})$ 的相应对数线性模型有以下形式:

$$\ln P(X=x, \mathbf{Y}=\mathbf{y}) = \beta + \beta_x^X + \sum_{l=1}^{L} \beta_{y_l}^{Y_l} + \sum_{l=1}^{L} \beta_{x,y_l}^{X,Y_l}$$

潜类模型包括一个主效应;潜变量和指标的单变量项,包括 X 和每个指标的双变量项。注意:由于局部独立假定,包括两个或多个显变量的项被略去了。

对数线性参数和条件响应概率之间的联系如下:

$$P(Y_l = y_l \mid X=x) = \frac{\exp(\beta_{y_l}^{Y_l} + \beta_{x,y_l}^{X,Y_l})}{\sum_{r=1}^{D_l} \exp(\beta_r^{Y_l} + \beta_{x,r}^{X,Y_l})}$$

它表明对数线性表达式相当于为每个条件响应概率设定一个logit模型。

如果希望加入约束条件,上述LC表达式就很重要了。虽然对概率的约束有时可以转换成对数线性参数的约束,反过来也可以,但是在许多情况下这种转换是不能进行的。

最大似然估计

令 I 为 L 维频数表的单元格总数(或者可能的回答组合数目),故而 $I = \prod_{l=1}^{L} D_l$,令 i 代表某一特定单元格,n_i 为单元格 i 中的观察频数,$P(\mathbf{Y}=\mathbf{y}_i)$ 为得到单元格 i 中的响应模式的概率。

LC模型的参数通常用最大似然法【Maximum Likelihood, ML】进行估计。被最大化的对数似然函数的核心为:

$$\ln L = \sum_{i=1}^{I} n_i \ln P(\mathbf{Y}=\mathbf{y}_i)$$

注意:只有观察频数非零的单元格才能进入对数似然函数中,过去几年发展出来的几个更有效的LC模型软件就是利用了这一特点。

LC模型估计中的一个问题是,模型参数可能是不可识别的,即使自由度【Degrees of Freedom】的数目大于或等于零。不可识别意味着有多组参数值可以使对数似然函数达到最大值,或者换而言之,不存在唯一的一组参数估计。正式的识别检验要通过信息矩阵,信息矩阵必须是正定的(positive definite)。另一种办法是用不同的起始值来

估计模型。除了局部解(local solution)外,可识别模型对于不同的起始值给出了相同的最终估计。

虽然对于 LC 模型的可识别性不存在普遍规则,但可提出一些最低要求和指出一些可能陷阱。对于一个非约束的(unrestricted)LC 分析来说,至少需要三个指标,但是如果这些指标都是二分的,最多两个潜类能够被识别。必须注意的是,四个二分变量是无法用来识别三个潜类的非约束模型的,即使自由度为正。但是,五个二分变量可以用来识别五个潜类的模型。通常可能通过约束某些模型参数来识别;例如,可以用约束条件 $P(Y_l=1 \mid X=1)=P(Y_l=2 \mid X=2)$ 的两个二分变量来识别两个潜类的模型。

与 LC 模型估计相关的第二个问题是多个局部最大值(local maxima)的存在。LC 模型的对数似然函数并不总是凹的,这意味着爬山算法(hill-climbing algorithms)可能会因不同的起始值而收敛于不同的最大值。我们通常要找的是全部最大值(global maximum)。因此,最好的处理办法是用不同的随机起始值去估计模型。通常不同的起始值会收敛于相同的最高对数似然值,这就被视为 ML 解。一些软件已经自动设置使用多组随机起始值来减少得到局部解的概率。

LC 模型的另一个问题在于边界解(boundary solution)的发生,即概率等于 0(或 1)或者对数线性参数等于负(或正)无穷。这会在估计算法【Algorithm】和局部解发生上引发问题,使得标准误差的计算和拟合优度的自由度计算变得复杂。边界解可以用加入约束条件或者考虑模型参数的其他先验信息的方法来避免。

解决 ML 估计问题的最常用方法是期望最大化算法(expectation-maximization,EM)和牛顿-拉普森算法(Newton-Raphson,NR)。EM 是一种对不完整数据进行最大似然估计的非常稳定的迭代方法。NR 是更快的程序,但是需要很好的起始值以便收敛。后一种方法要用到对数似然函数的二阶导数矩阵,在计算模型参数的标准误差时也需要这一矩阵。

模型选择问题

一个估计出的 LC 模型的拟合优度通常用皮尔逊卡方统计量或者似然比卡方统计量来进行检验(参见分类数据分析【Categorical Data Analysis】)。后者的定义为:

$$L^2 = 2\sum_{i=1}^{I} n_i \ln \frac{n_i}{N \cdot P(\mathbf{Y}=\mathbf{y}_i)}$$

其中 N 代表总样本量。与对数线性分析中一样,自由度(df)数目等于频数表中的单元格数目减去 1,即 $\prod_{l=1}^{L} D_l - 1$,再减去独立参数数目。在非约束的 LC 模型中,自由度为:

$$df = \prod_{l=1}^{L} D_l - C \cdot \left[1 + \sum_{l=1} (D_l - 1)\right]$$

虽然对 10,20 或 50 个指标来估计 LC 模型没有什么问题,但是在上述情况下频数表会变得非常稀疏,因此渐进的 p 值也不再可靠。对这个问题的一个很好但有点费时的解决方法是用参数性的自助法(parametric bootstrapping)来估计 p 值。另一种方法是在低阶的边缘表中(如在二维边缘表中)评估模型拟合。

通过将 L^2 和 df 相减来比较有 C 和 $C+1$ 个潜类的模型是无效的,因为这种条件检验并不服从渐进的卡方分布。这意味着需要用其他方法来比较有着不同潜类数目的模型。一种通行的方法是使用信息准则,如贝叶斯信息量准则(BIC)或者赤池信息准则

(AIC);另一种更具描述性的方法是测量用 C 个潜类的模型能够解释总关联的比例,$[L^2(1) - L^2(C)]/L^2(1)$,其中 1 个潜类的模型(独立模型)的 $L^2(1)$ 被用来表示 L 维列联表中的总关联度。

我们通常不只对拟合优度感兴趣,而且对模型分类准则的表现感兴趣(参见公式(3))。在模型分类下分类出现错误的估计比例等于:

$$E = \sum_{i=1}^{I} \frac{n_i}{N}\{1 - \max[P(X = x \mid \mathbf{Y} = \mathbf{y}_i)]\}$$

这个数字可以与基于非条件概率 $P(X=x)$ 的分类错误比例进行比较,从而得到以下误差消减测度 λ:

$$\lambda = 1 - \frac{E}{\max[P(X = x)]}$$

这个测度与 R^2 有类似之处,它越接近 1,模型的分类表现就越好。

LC 模型对类别指标的扩展

基本的 LC 模型已经有几个扩展形式。最重要的扩展之一是包括能够描述(或预测)潜变量 X 的协变量或分组变量。办法是为属于 LC x 的概率设定一个多项 logit 模型,即

$$P(X = x \mid \mathbf{Z} = \mathbf{z}) = \frac{\exp(\gamma_x^X + \sum_{k=1}^{K} \gamma_x^{X,Z_k} \cdot z_k)}{\sum_{r=1}^{C} \exp(\gamma_r^X + \sum_{k=1}^{K} \gamma_r^{X,Z_k} \cdot z_k)}$$

其中的 z_k 代表协变量 k 的一个取值。

另一个重要的扩展与类别的定序信息有关。在对数线性的 LC 框架中,可以通过两个变量项 β_{x,y_l}^{X,Y_l} 的关联模型【Association Model】结构来加入定序约束。例如,如果 Y_l 是定序指标,我们可以加入约束 $\beta_{x,y_l}^{X,Y_l} = \beta_x^{X,Y_l} \cdot y_l$。同样也可以对潜变量施加类似约束(Heinen, 1996)。

如果 C 个潜类的模型不能拟合数据,局部独立假定对于某一对或更多对指标来说就不能成立。在 LC 分析中,常用的建模策略是增加潜类的数量,直到局部独立假定成立。已经有两种模型扩展可以使我们采用其他策略。一种方法是不去增加潜类数量,而是通过包括某些指标间的直接效应来放松局部独立假定,这是对数线性 LC 模型的直接扩展;另一种方法是增加潜变量的数目而不是潜类的数目。如果指标测量的是多个维度,那么这种所谓的 LC 因子分析就特别有用(Magidson & Vermunt, 2001)。

其他重要的扩展包括分析纵贯数据(参见潜马尔科夫模型【Latent Markov Model】),以及部分可观察指标。哈格纳尔斯(Hagenaars, 1990)和费蒙特(Vermunt, 1997)提出了最为一般的类别数据的结构方程模型,前面讨论过的所有模型都是这种一般模型的特例。

LC 模型的其他类型

到目前为止,我们讨论的是对类别数据的 LC 模型。但是 LC 分析的基本理念,即统计模型的参数在不可观察的小组间有所不同,也可应用到其他测度类型的变量上来。特别是有三种重要的 LC 模型或有限混合模型的应用是在类别数据分析框架之外的,即连续变量的聚类、密度估计、随机效应模型。

在过去 10 年间,人们又开始对用 LC 分析对连续变量进行聚类分析感兴趣。LC 分析可以视为传统的非层级聚类分析(如 K-means方法)的一种基于概率或基于模型的变体。已经可以证明,这样一种基于 LC

的聚类程序优于更为特殊化的(ad hoc)传统方法。这种方法也被称为潜相模型【Latent Profile Model】、混合模型聚类(mixture-model clustering)、基于模型的聚类(model-based clustering)、潜鉴别分析(latent discriminant analysis)、潜类聚类(LC clustering)。这种模型的基本公式与公式(1)类似,即

$$f(\mathbf{Y}=\mathbf{y})=\sum_{x=1}^{C}P(X=x)f(\mathbf{Y}=\mathbf{y}\mid X=x)$$

在上述略为一般化的公式中,概率$P(\cdots)$被替换为密度$f(\cdots)$。对于连续变量来说,分类别的密度$f(\mathbf{Y}=\mathbf{y}\mid X=x)$通常被假定服从(有约束的)多元正态分布,其中每个 LC 有自己的均值向量和协方差矩阵。注意,这是用简单密度的有限混合来进行密度估计的更为一般的方法的特例。

LC 分析的另一种重要应用是作为一种非参数随机效应模型【Nonparametric Random-Effect Model】。这种应用背后的理念是回归模型的参数可以在不可观察的分组间存在差异。对于这类常被称为 LC 回归分析的方法来说,LC 变量起到了调节变量的作用。这种方法类似于重复测量的回归模型,或者双水平数据的回归模型,区别在于对随机系数的分布不需要进行任何假定。

软件

1977 年,克里弗德·克劳格写出了第一个 LC 程序 MLLSA,它只能用于分析相对少的定类变量。如今的程序可以处理的变量要多得多,而且可以处理很多测度类型的变量。例如,LEM 程序(Vermunt, 1997)提供的命令语言可以用来设定许多类别数据模型,包括 LC 模型。Mplus 是基于命令语言的结构方程建模软件,可以用来执行某些 LC 模型,但是不能用于定类指标。与这些命令语言程序不同,Latent GOLD 是专门用于 LC 分析的程序,有着和 SPSS 类似的用户界面。它可以执行 LC 模型中最为重要的类型,可以处理不同测度类型的变量,可以将基本模型进行扩展以包括协变量、局部相依、多个潜变量、部分可观察指标等。

——Jeroen K. Vermunt
Jay Magidson
(高勇译校)

参考文献

Goodman, L.A. (1974). The analysis of systems of qualitative variables when some of the variables are unobservable: I. A modified latent structure approach. *American Journal of Sociology*, *79*, 1179-1259.

Haberman, S.J. (1979). *Analysis of qualitative data: Vol. 2. New developments*. New York: Academic Press.

Hagenaars, J. A. (1990). *Categorical longitudinal data: loglinear analysis of panel, trend and cohort data*. Newbury Park, CA: Sage.

Hagenaars, J. A., & McCutcheon, A. L. (2002). *Applied latent class analysis*. Cambridge, UK: Cambridge University Press.

Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences.* Thousand Oaks, CA: Sage.

Lazarsfeld, P.F. (1950). The logical and mathematical foundation of latent structure analysis & the interpretation and mathematical foundation of latent structure analysis. In S. A. Stouffer (Ed.),

Measurement and prediction (pp. 362-472). Princeton, NJ: Princeton University Press.
Magidson, J., & Vermunt, J.K. (2001). Latent class factor and cluster models, bi-plots and related graphical displays. *Sociological Methodology*, *31*, 223-264.
Vermunt, J. K. (1997). *Log-linear models for event histories*. Thousand Oaks, CA: Sage.

潜马尔科夫模型(Latent Markov Model)

潜马尔科夫模型(LMM)可以视为潜类【Latent Class】模型应用于纵贯数据分析时的扩展,也可以视为离散时间马尔科夫链【Markov Chain】模型处理可观察变量存在测量误差时的扩展。威金斯在1955年首次引入这一模型,也被称为“潜转换模型(latent transition model)”或者“隐含马尔科夫模型(hidden Markov model)”。LMM可以用来区分个体行为中真实的系统性变迁与由于测量误差或者其他随机因素而造成的虚假变化。

假定我们感兴趣的一个类别变量在 T 个时点被加以测量,Y_t 表示在时点 t 的测量结果,$1 \leq t \leq T$。例如,每6个月测量一次被访者的党派倾向。令 D 表示每个 Y_t 的类别数,令 y_t 表示某一具体类别,$1 \leq y_t \leq D$。令 X_t 表示因时而变的潜变量,C 表示 X_t 的类别数,x_t 表示在时点 t 上的具体潜类别,$1 \leq x_t \leq C$。相应的LMM有以下形式:

$$P(\mathbf{Y} = \mathbf{y}) = \sum_{x} P(X_1 = x_1) \times \prod_{t=2}^{T} P(X_t = x_t \mid X_{t-1} = x_{t-1}) \times \prod_{t=1}^{T} P(Y_t = y_t \mid X_t = x_t)$$

模型有两个基本假定:潜变量的转换结构具有一阶马尔科夫链的形式;每个因时而变的可观察变量只取决于相应的潜变量。为了识别和结果的简化,通常会假定误差构成是不随时间变化的,即对于 $2 \leq t \leq T$,有:

$$P(Y_t = y_t \mid X_t = x_t) = P(Y_{t-1} = y_t \mid X_{t-1} = x_t)$$

如果不施加其他约束条件,至少需要三个时点才能识别LMM。对于潜转换概率的通常约束条件包括时间同质性和零约束。

用一个假想示例可以说明研究变迁时测量误差的巨大影响,这个示例中 $T=3$,$C=D=2$。假定 $P(X_1 = 1) = 0.80$,$P(X_t = 2 \mid X_{t-1} = 1) = P(X_t = 2 \mid X_{t-1} = 1) = 0.10$,$P(Y_t = 1 \mid X_t = 1) = P(Y_t = 2 \mid X_t = 2) = 0.20$。如果我们对于假想表格估计一个静态的显一阶马尔科夫模型,会发现在第一个时点位于第一种状态的概率为0.68,两种状态发生转换的概率分别为0.29和0.48。这就是不考虑测量误差时遇到的典型偏差:较小组别的规模被夸大了,变迁规模被夸大了,较小组别的变迁程度会显得比较大组别的变迁程度更大。

显而易见,可以将上述单指标LMM扩展到多个指标的情况中。另一种自然的扩展是引入协变量或分组变量来解释个体在初始状态和转换概率上的差异。独立分类误差【Independent Classification Error, ICE】假定可以放松,即可以让不同时点的指标之

间存在直接效应。此外,有人提出了 LMM 的混合衍生模型,如有着迁移留守【Mover-Stayer】结构的模型。在社会科学中,LMM 被认为是一种类别数据分析工具。但是,如同标准的潜类分析【Latent Class Analysis】一样,这些模型可以很容易地扩展到其他测量尺度中。

PANMARK 程序可以用于估计较为标准的 LMM。LEM 程序也可以处理更为扩展的模型,如包括协变量和指标间直接效应的模型。

——Jeroen K. Vermunt

(高勇译校)

参考文献

Langeheine, R., & Van de Pol, F. (1994). Discrete-time mixed Markov latent class models. In A. Dale & R. B. Davies (Eds.), *Analyzing social and political change: A casebook of methods* (pp. 171-197). London: Sage.

MacDonald, I. L., & Zucchini, W. (1997). *Hidden Markov models and other types of models for discrete-valued time series*. London: Chapman & Hall.

Vermunt, J. K. (1997). *Log-linear models for event histories*. Thousand Oaks, CA: Sage.

Wiggins, L. M. (1973). *Panel analysis*. Amsterdam: Elsevier.

潜相模型(Latent Profile Model)

潜相模型是一种潜变量【Latent Variable】模型,其中的潜变量为类别变量,显变量为连续变量。拉扎斯菲尔德和亨利于 1968 年提出了这种模型。与此同时,戴(Day,1969)和沃尔夫(Wolfe,1970)也提出了非常类似的模型,虽然名称有所不同。

过去 10 年间,人们重新对这种潜变量模型产生了兴趣,特别是作为一种聚类分析【Cluster Analysis】的工具。潜相模型可以看成传统的非层次聚类分析程序(如 K 均值方法)的概率性衍生或基于模型的衍生。人们已经证明,基于模型的聚类程序优于更为特殊化的(ad hoc)传统方法。需要注意的是,只有某些作者使用"潜相模型(latent profile model)"这一名称。更为常用的名称包括正态成分混合(mixture of normal components)、混合模型聚类(mixture model clustering)、基于模型的聚类(model-based clustering)、潜判别分析(latent discriminant analysis)、潜类聚类(latent class clustering)。

潜相模型在社会科学中的可能应用包括建立类型(typology)和构建诊断工具。社会学家可用它来基于一组社会经济指标和政治指标对国家进行分类。心理学家可用这种方法综合多个测试得分以得到一个诊断工具。

与潜类分析【Latent Class Analysis】一样,潜相分析假定总体中包括了 C 个不可观察的小组,这些小组被称为潜相、潜类或者混合成分。由于指标是连续变量,假定其条件分布服从正态分布是最自然的。在非约束性的多元正态分布下,可以得到最一般的模型,即

$$f(\mathbf{y}) = \sum_{x=1}^{C} P(x) f(\mathbf{y} \mid \boldsymbol{\mu}_x, \Sigma_x)$$

上述公式说明,L 个指标的联合密度 $f(\mathbf{y})$ 是各个潜类的密度(class-specific densities)的混合。每个潜类 x 有其自身的均值向量 $\boldsymbol{\mu}_x$ 和协方差矩阵 Σ_x。在每个成分中的人数比例表示为 $P(x)$。注意,模型

结构与二次判别分析(quadratic discriminant analysis)很类似,当然根本差异在于这里的类别(组)是未知的。

对协方差矩阵 Σ_x 进行约束,就可以得到几种特例。常用的约束包括各个类别的协方差矩阵相同;协方差矩阵为对角矩阵;协方差矩阵相同且为对角矩阵。协方差矩阵相同的假定类似于线性判别分析的基本假定。将协方差矩阵设定为对角矩阵,相当于局部独立【Local Independence】假定。该模型可以表达为以下公式:

$$f(\mathbf{y}) = \sum_{x=1}^{C} P(x) \prod_{l=1}^{L} f(y_l \mid \mu_{lx}, \sigma_{lx}^2)$$

假定局部独立且误差方差相同,即 $\sigma_{lx}^2 = \sigma_l^2$,这种设定类似于 K 均值聚类。但是,在大多数情况下,这种局部独立的设定都太过严格了。

最近研究者提出了几种构建各个类别的协方差矩阵的方法。一种相对简单的方法是块对角矩阵(block-diagonal matrices),它是完全协方差矩阵和对角协方差矩阵之间的折中方案。这相对某些指标放松了局部独立假定。更为复杂的方法要使用主成分分析和因子分析来分解协方差矩阵,以及对 Σ_x 进行结构方程模型类型的约束。

另一个最近的进展是综合了连续指标和类别指标的模型。它实际上是经典的潜类模型和潜相模型的综合。最近的扩展是包括协变量来预测类别归属。

可以用来估计潜相模型的软件包包括 EMMIX,Mclust,Mplus 及潜相 GOLD。

——Jeroen K. Vermunt

(高勇译校)

参考文献

Day, N. E. (1969). Estimating the components of a mixture of two normal distributions. *Biometrika*, *56*, 463-474.

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.

McLachlan, G. J., & Basford, K. E. (1988). *Mixture models: Inference and application to clustering*. New York: Marcel Dekker.

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89-106). Cambridge, UK: Cambridge University Press.

Wolfe, J. H. (1970). Pattern clustering by multivariate mixtures. *Multivariate Behavioral Research*, *5*, 329-350.

潜特质模型(Latent Trait Models)

潜特质模型是一种描述在某一题项上得到特定得分(如正确/不正确,或某一定序量度上的得分)的概率和解释这一得分的潜变量【Latent Variable】或潜在特质之间关系的测量模型。对于回答某一特定题项的个体来说,这个概率取决于他/她的潜在特质和题项的性质,如它的难度和鉴别度。潜质模型这一术语是题项反应模型的同义词,但如今并不像后者那样常用。

——Klaas Sijtsma

(高勇译校)

* 也可参见项目反应理论【Item Response Theory】。

潜变量(Latent Variable)

社会科学家感兴趣的许多构念都不能直接观察到,例如偏好、态度、行为意图、人格特质。此类构念只能用可观察指标的方式来进行间接测量,如与态度或偏好相关的问卷题项就被设计出来以得到回答。许多种度量技术被发展出来,以便从指标中获取不可观察的构念的信息。这些度量方法中的重要一类就是通过潜变量模型构建出来的。

一个潜变量模型是一个可能非线性的路径分析【Path Analysis】模型或图形模型。除了显变量外,模型包括了一个或多个不可观察的潜变量,代表着我们感兴趣的构念。两个假定定义了回答背后的因果机制。首先我们假定,对指标的回答是个体在潜变量上位置的结果。第二个假定是,在控制了潜变量之后,显变量之间不再有任何共同因素,这通常被称为局部独立【Local Independence】原理。

其余两个假定与潜变量分布和显变量分布有关。取决于这些假定,我们可以得到不同类型的潜变量模型。依据巴塞洛缪和诺特(Bartholomew & Knott,1999)的划分,存在如下四种主要类型:因子分析【Factor Analysis, FA】、潜质分析(latent trait analysis, LTA)、潜相分析(latent profile analysis, LPA)和潜类分析【Latent Class Analysis,LCA】,参见表1。

表1

显变量	潜变量	
	连续变量	类别变量
连续变量	因子分析	潜相分析
类别变量	潜质分析	潜类分析

在FA和LTA中,潜变量被处理为连续正态分布变量。而在LPA和LCA中,潜变量是离散的,因此被假定出自多类别分布。在FA和LPA中的显变量是连续的。在大多数情况下,它们在潜变量给定条件下的条件分布被假定为正态的。在LTA和LCA中,指标是二分的、定序的或者类别变量,它们的条件分布被假定为二项分布或多项分布。

巴塞洛缪和诺特(Bartholomew & Knott,1999)的分类中,连续潜变量和离散潜变量之间的区分更为根本。研究者必须确定潜变量被处理为连续的还是离散的更好。但是,如海内恩(Heinen,1996)所示,连续潜变量的分布可以用离散分布来近似。这表明,连续潜变量和离散潜变量之间的区分并不像人们想象的那样重要。

连续指标和离散指标之间的区分不是根本性的。指标的条件分布设定自然地取决于其量度类型。潜变量模型最近的发展已经可以允许每个指标有不同的分布形式。例如,连续变量可以服从正态分布、学生分布、对数正态分布、伽马分布、指数分布;二分变量可以服从二项分布;定序和类别变量可以服从多项分布;计数变量可以服从泊松分布、二项分布或负二项分布。根据潜变量被处理为连续的还是离散的,我们可以得到LTA和LCA的推广形式。

——Jeroen K. Vermunt

Jay Magidson

(高勇译校)

参考文献

Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis*. London: Arnold.

Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences*. Thousand Oaks, CA: Sage.

拉丁方(Latin Square)

拉丁方是字母或有序符号的方阵,其中每个字母或符号在每一行和每一列中均出现且只出现一次。下面是2×2,3×3和4×4的拉丁方示例。

A	*B*	*A*	*B*	*C*	*A*	*B*	*C*	*D*
B	*A*	*B*	*C*	*A*	*B*	*A*	*D*	*C*
		C	*A*	*B*	*C*	*D*	*B*	*A*
					D	*C*	*A*	*B*

“拉丁方”这一名称是由著名的瑞士数学家欧拉(Léonard Euler,1707—1783)赋予这种方阵的,他研究了这种方阵,并使用了拉丁字母。

上表中的三个方阵被称为标准方阵,因为其第一行和第一列均按字母表顺序排列。如果某一方阵的行与另一方阵的列完全相同,这两个拉丁方就是共轭的。例如,一个5×5 拉丁方及其共轭如下所示:

A	*B*	*C*	*D*	*E*	*A*	*B*	*C*	*D*	*E*
B	*A*	*D*	*E*	*C*	*B*	*A*	*E*	*C*	*D*
C	*E*	*A*	*B*	*D*	*C*	*D*	*A*	*E*	*B*
D	*C*	*E*	*A*	*B*	*D*	*E*	*B*	*A*	*C*
E	*D*	*B*	*C*	*A*	*E*	*C*	*D*	*B*	*A*

如果一个拉丁方行列互换之后得到相同的拉丁方,这个拉丁方就是自共轭的。上面的 2×2,3×3 和 4×4 拉丁方都是自共轭的。

一个拉丁方中的 p 个不同字母可以进行重新组合,得到 $p!\ (p-1)!$ 个拉丁方,包括原初的拉丁方。例如,存在 $3!\ (3-1)! = 12$ 个 3×3 的拉丁方。费希尔和叶芝(Fisher & Yates,1934)、诺顿(Norton,1939)和萨德(Sade,1951)已列举出了从 2 × 2 到 7×7 的所有拉丁方。

拉丁方是有趣的数学现象,在设计实验中也很有用。拉丁方设计可以使研究者在分离出两个各自有 p 个水平的冗余变量效应的同时进行 p 个总体的均值是否相同的假设检验。冗余变量是实验中不想要的变异来源。通过指定一个冗余变量为拉丁方的行,另一个冗余变量为拉丁方的列,两个冗余变量的效应就可以得到分离。处理的 p 个水平被指定为方阵的拉丁字母。分离出两个冗余变量通常会降低均方误差【Mean Square Error】,导致统计效力增加。

考虑如下例子。研究者想要比较三种饮食方案在帮助肥胖男童减肥方面的效率。自变量是三种饮食方案;因变量是坚持饮食方案三个月后减掉的体重数。我们不用字母 *A*,*B* 和 *C* 来代表不同饮食方案,而是遵循现代用法,用小写字母 a 和下标来代表不同饮食方案:a_1,a_2和 a_3。研究者认为,减肥的难度受到男童超重数量和超重基因倾向的影响。后一个变量的近似测量可通过询问男童的父母在青少年时期是否超重来获取:c_1表示父母双方均不超重;c_2表示父母中有一方超重;c_3表示父母均超重。

	c_1	c_2	c_3
b_1	a_1	a_2	a_3
b_2	a_2	a_3	a_1
b_3	a_3	a_1	a_2

通过分离出两个冗余变量的影响,研究者可以获得对自变量的更具效力的检验。

拉丁方设计的一个重要假定是在冗余变量和自变量之间不存在交互【Interactions】作用。如果此假定不满足,统计检验将会是有偏误的。

农学家弗朗索瓦·克雷泰·帕吕埃尔(Francois Cretté de Palluel,1741—1798)被认为在1788年进行了第一次基于拉丁方的实验。他的实验包括4个品种的绵羊,用4种饲养方案饲养,在连续4个月的20日进行屠宰。推广基于拉丁方的实验设计的使用最有功者是卓越的英国统计学家罗纳德·费希尔(Ronald A. Fisher,1890—1962),他在伦敦北部的Rothamsted实验站工作。费希尔表明,拉丁方及其组合可以用来构建许多复杂的实验设计(关于这些设计的描述,请参见Kirk,1995, Chaps.8,13,14)。

——Roger E. Kirk

(高勇译校)

参考文献

Fisher, R. A., & Yates, F. (1934). The six by six Latin squares. *Proceedings of the Cambridge Philosophical Society*, *30*, 492-507.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Norton, H. W. (1939). The 7 × 7 squares. *Annals of Eugenics*, *9*, 269-307.

Sade, A. (1951). An omission in Norton's list of 7 × 7 squares. *Annals of Mathematical Statistics*, *22*, 306-307.

平均数定律(Law of Averages)

随着试验次数的增加,随机误差【Random Error】会成比例地减小。因此,按照平均数定律,在足够多次试验后,结果会更接近期望值。以掷硬币为例,掷100次,正面朝上的期望值为50次,但是存在偶然因素,结果可能与此有差距。比如说,由于偶然因素,结果与期望值差5次。如果掷硬币1 000次,正面朝上的期望值仍然是一半。但是由于存在偶然因素,结果与期望值有差距,比如差8次。误差的绝对数目增加了,从5增加到8,但是误差的比例减小了,从5%减小到8‰。总而言之,平均数定律讲的是,随着掷硬币次数增加,随机误差的总比例会减小。因此,在掷硬币的分布中,正面朝上的比例会更接近期望值50%。

——Michael S. Lewis-Beck

(高勇译校)

大数定律(Law of Large Number)

大数定律是数理统计学中最重要的结果之一,它描述的是随机变量的观察行为(即从基于随机抽样【Sample】的调查中得到的观察)。大数定律关心的是一个变量的期望值【Expected Value】与从大样本中计算出的平均数【Average】的差异。该定律说明了这种差异不可能很大。

在定律的最简单形式中,存在一个事件

A(如观察到被访者具有某一特定属性),其概率【Probability】为 $P(A)$(概率等于总体中具有这种属性的人的比例),在进行了 n 次观察后得到其相对频率【Relative Frequency】为 $r(n,A)$。大数定律说的是,只要 n 足够大,$r(n,A)$ 将会以任意大的概率,以任意程度接近 $P(A)$。更确切地说,对于任意小的两个值 a 和 b,都存在一个临界值,使得对于所有大于这个临界值的样本量 n 而言,下式成立:

$$P(|P(A) - r(n,A)| < a) > 1 - b$$

事件概率与相对频率的绝对差值小于 a 的概率会大于 $1-b$,只要样本量 n 足够大。但是绝对差值大于 a 的概率仍然存在,只不过很小(小于 b)。

这个结果的重要性体现在两个方面。首先,频率学派认为只有相对频率具有稳定性时(即从大样本中计算出来的相对频率会彼此接近时),谈论概率才有意义,并且认为相对频率就是不可直接观察的概率的外在显示。结果表明了这一性质,这就显示了概率的数理理论恰当地包含了这一基本假定。其次,大数定律也是数理统计学的基本结论。在此背景下,它判断,只要样本量足够大,估计量(事件的相对频率)就会非常接近真值(事件的概率)。

前面的结果并没有说明样本量要多大才能保证相对频率接近概率。从这个方面来看,下面的表述更为有用。令 $E(X)$ 和 $D(X)$ 是一个变量的期望值与标准差【Standard Deviation】。那么,对于任意正值 c,下式成立:

$$P[|X - E(X)| > cD(X)] < \frac{1}{c^2}$$

上述结果用标准差单位来测量观察值(即相对频率)和期望值(即概率)之间的差值。差值超过 c 倍的标准值的概率小于 $1/c^2$。这一结论适用于任何从调查中观察到的量。

例如,在一个城镇中用调查来估计支持某一动议的比例。如果样本量为 100,可以对估计的可能结果作出下述界定:从二项【Binomial】分布可知,标准差不会超过 5%,因此,调查结果偏离真值超过 10%(即 2 倍标准差)的概率不会超过 1/4=0.25。

——Tamás Rudas

(高勇译校)

参考文献

Feller, W. (1968). *An introduction to probability theory and its applications* (3rd ed.). New York: John Wiley.

社会科学中的法则(Laws in Social Science)

社会科学中是否存在法则,社会科学家在这个问题上会按学科立场和认识立场各自站队。宏观经济学传统上就倚重于法则,至少是类似法则的陈述(如"供求定律"),而多数历史学家会对历史法则的存在表示怀疑。这些立场(当然也存在反例)反映了这些学科的认识论基础。

拒斥法则存在的认识论基础可以概括为如下观点:由于个体意识与自由意识,因此存在太多变数,以至于社会世界中不可能存在与类似法则一样的规律性。这一论证结论从根本上讲是正确的,但是它常常预设了一个过于简单的"法则"概念。在通常理解的科学法则概念中,不允许在预测现象行

为时有例外。物理学的经典法则是最常被提及的,如重力法则、力学法则等。但是,多数科学法则有着非常不同的性质,甚至经典法则也会展现出一些局部的例外,在特定情况下(在人类经验之外)会完全失效(Nagel, 1979, Chaps.4,5)。

法则表达的可以是基本的理论原则(经济学中就是如此),也可以是数学公理,或者通过实验发现的不变性,或者统计规律。虽然理论通常被视为是对规律的更侧重于推理的陈述,但是有时法则与理论之间的区别只是出于语言或社会的便利。因此,虽然牛顿法则和爱因斯坦理论的可靠程度相近,但是人们还是称呼一个是法则,一个是理论(Williams, 2000, p.33)。如果能够接受理论证明是可能的,从而也接受社会世界中存在规律性,那么这是否意味着有些法则能够解释事情并非偶然存在呢?在某些情况下,经济行为和人口行为中的规律也许与气体规律一样是可预测的。至少统计法则看起来可能如此。但是,沿用气体的类比,虽然某些当地条件(如温度或气压的变化)会使得气体分布在统计上有所波动,但是人口法则和经济法则的例外更多要取决于独特的文化情境或历史情境(Kincaid, 1994)。因此,在物理法则的例外中存在规律性,但是社会法则表明这种规律性需要诉诸其他普遍原则,如理性。

比是否存在社会法则更重要的是,有充分证据表明社会世界中存在着并非偶然的规律,即使此类规律在文化和历史适用程度上存在差异。

——Malcolm Williams

(高勇译校)

参考文献

Kincaid, H. (1994). Defending laws in the social sciences. In M. Martin & L. McIntyre (Eds.), *Readings in the philosophy of social science* (pp. 111-130). Cambridge: MIT Press.

Nagel, E. (1979). *The structure of science: Problems in the logic of scientific explanation*. Indianapolis, IN: Hackett.

Williams, M. (2000). *Science and social science: An introduction*. London: Routledge Kegan Paul.

最小二乘(Least Squares)

最小二乘是一种非常流行的用于计算参数估计和拟合数据的技术。它是现代统计学中历史最悠久的技术之一,首次出现于法国数学家勒让德在 1805 年发表的一本现已成为经典的备忘录中。著名的德国数学家高斯在 1809 年发表了另一本备忘录,其中他提及早在 1795 年他就已经发现并使用了这一方法。由此导致了一场激烈的发现优先权争议(有点像莱布尼茨和牛顿对微积分发明的争议),但是这并没有妨碍这种技术很快流行起来。高尔顿 1886 年在其著作中使用它研究了身高体重的遗传性,这奠定了相关【Correlation】分析和回归【Regression】分析的基础,并为回归分析命名。在统计学的早期发展中贡献巨大的皮尔逊和费希尔,都在不同的背景下使用并发展了这一技术(对皮尔逊而言是因子分析【Factor Analysis】;对费希尔而言则是实验设计【Experiment Design】)。

如今,最小二乘已广泛使用于发现或估计参数数值,以此来对数据进行拟合或者对估计量的统计特质进行把握。它有如下几

种变化形式：最简单的形式被称为普通最小二乘【Ordinary Least Squares，OLS】；更为复杂的形式被称为加权最小二乘【Weighted Least Squares，WLS】，它通常比 OLS 表现得更好，因为它能够在最后结果中调整每个观察的重要性；最小二乘的最新变化形式是交替最小二乘【Alternating Least Squares，ALS】和偏最小二乘【Partial Least Squares，PLS】。

函数拟合实例：回归

OLS 的最早应用（也是最常见的应用）就是线性回归【Linear Regression】，它要解决的问题是找到能够最佳拟合数据的一条直线或曲线。在标准公式中，N 对观察值 $\{Y_i, X_i\}$ 用来找到一个自变量（X）取值到因变量（Y）取值的函数。在单变量线性函数的前提下，预测由下述等式给出：

$$\hat{Y} = a + bX \tag{1}$$

这个等式包括两个自由参数，分别界定了回归线的截距（a）和斜率（b）。最小二乘将这些参数估计定义为能够使测量与模型（即预测值）之间差异的平方和最小的值（由此得名最小二乘）。这相当于最小化如下表达式：

$$\varepsilon = \sum_i (Y_i - \hat{Y}_i)^2 = \sum_i [Y_i - (a + bX_i)]^2 \tag{2}$$

其中的 ε 代表误差项，这就是要最小化的量。这可通过微积分中的标准技术来得到，即二次方程（有平方项的方程）在其导数为零时可以得到其最小值。分别对 a 和 b 求 ε 的导数，将其设定为零，然后就可得到被称为正规方程（normal equation）的式子：

$$\frac{\partial \varepsilon}{\partial a} = 2Na + 2b \sum X_i - 2 \sum Y_i = 0 \tag{3}$$

$$\frac{\partial \varepsilon}{\partial b} = 2b \sum X_i^2 + 2a \sum X_i - 2 \sum Y_i X_i = 0 \tag{4}$$

解上述两个方程，可以得到 a 和 b 的最小二乘估计为：

$$a = M_Y - bM_X \tag{5}$$

其中，M_Y 和 M_X 分别代表 X 和 Y 的均值，并且：

$$b = \frac{\sum (Y_i - M_Y)(X_i - M_X)}{\sum (X_i - M_X)^2} \tag{6}$$

OLS 可以扩展到自变量多于一个和函数形式是非线性的情况（要使用到矩阵代数【Matrix Algebra】）。

最小二乘的几何意义

OLS 可以在几何框架中解释为数据向量在自变量定义空间上的正交投射。投射之所以是正交的，是因为预测值和实际值是不相关的。这可以在图 1 中来说明，图中描绘了两个自变量（向量 $\mathbf{x}_1$ 和 $\mathbf{x}_2$）和数据向量（$\mathbf{y}$），误差向量（$\mathbf{y} - \hat{\mathbf{y}}$）正交于位于两个自变量定义的子空间中的最小二乘估计（$\hat{\mathbf{y}}$）。

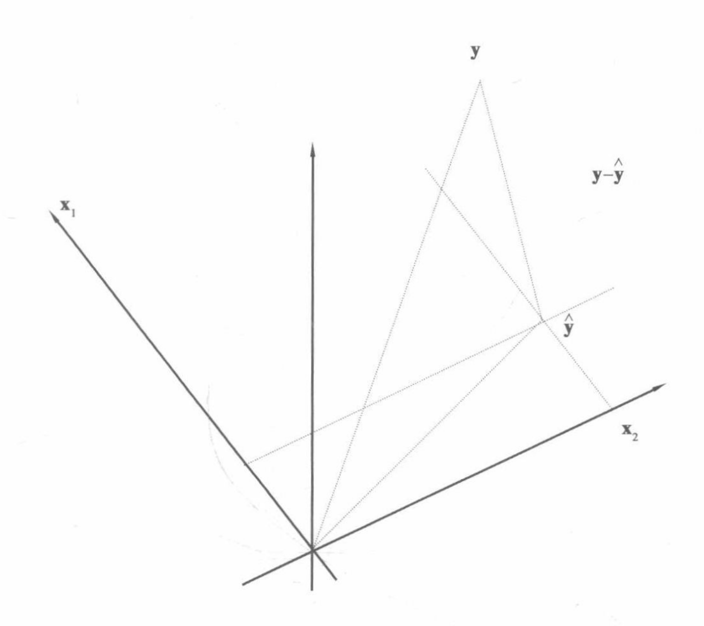


图 1 数据的最小二乘估计是数据向量在自变量子空间上的正交投射

最小二乘估计的最优性

OLS 估计有一些较强的统计性质。特别地,当以下条件成立时:(a)获得的数据是从定义清晰的总体【Population】中抽取的随机样本【Random Sample】;(b)总体模型是线性的;(c)误差项期望值为零;(d)自变量是线性独立的;(e)误差项服从正态分布且与自变量无关(即所谓的同方差性假设),此时 OLS 估计是最佳线性无偏估计【Best Linear Unbiased Estimator, BLUE】,通常用其首字母缩写 BLUE 表示(上述五个条件和证明被称为高斯-马尔科夫条件和定理)。此外,如果高斯-马尔科夫条件成立,OLS 估计同时也是最大似然【Maximum Likelihood】估计。

加权二乘估计

OLS 估计的最优性严重依赖于同方差性【Homoskedasticity】假定。如果数据来自不同的子总体,可以获得这些子总体中彼此独立的误差项的方差估计值,那么可以用加权最小二乘(WLS)来得到比 OLS 更好的估计量,WLS 也被称为广义最小二乘【Generalized Least Squares, GLS】。其基本思路是赋予每个观察值一个反映其测量不确定性的权重。一般而言,赋予第 i 个观察值的权重 w_i 是这个观察值的方差 σ_i^2 的函数。最直接的一种加权方案就是 $w_i = \sigma_i^{-1}$(也可以提出更为复杂的其他加权方案)。对于线性回归的例子来说,WLS 就是要发现能使下列式子最小化的 a 和 b 的值:

$$\varepsilon_\omega = \sum_i \omega_i (Y_i - \hat{Y}_i)^2 = \sum_i \omega_i [Y_i - (a + bX_i)]^2 \qquad (7)$$

迭代方法:梯度下降法

在用 OLS 或 WLS 来估计非线性函数的参数时,使用导数的标准方法并不是总能行得通。这时通常就要用到迭代方法。这些方法会逐步搜寻估计量的最佳取值。通常,它们会在每一步中使用函数的一个线性近似,然后再通过不断修正来改善近似。其中

涉及的技术被称为梯度下降(gradient descent)和高斯-牛顿近似(Gauss-Newton approximations)。它们对应于数值分析中的非线性最小二乘近似和统计学中的非线性回归。神经网络【Neural Network】是这些技术中最新近流行起来的应用。

——Hervé Abdi

(高勇译校)

参考文献

Bates, D. M., & Watts, D. G. (1988). *Nonlinear regression analysis and its applications.* New York: John Wiley.

Greene, W. H. (2002). *Econometric analysis.* New York: Prentice Hall.

Nocedal, J., & Wright, S. (1999). *Numerical optimization.* New York: Springer-Verlag.

Plackett, R. L. (1972). The discovery of the method of least squares. *Biometrika*, *59*, 239-251.

Seal, H. L. (1967). The historical development of the Gauss linear model. *Biometrika*, *54*, 1-23.

最小二乘原理(Least Squares Principle)

参见回归【Regression】。

离开田野(Leaving the Field)

在民族志文献中,对进入田野有大量论述,但是很少有人关注其必然结果:离开田野。离开田野也被称为“脱离(disengagement)”,就是退出田野场景和脱离研究行动培育起来的社会关系的过程。

离开田野会受到以下因素的影响:田野工作者进入的方式,他们在田野中培育的社会关系的强度,他们作出的研究交易,对于信息提供人而言,他们变成了什么样的人——或者研究者和被研究者是如何参与到罗奇福德(Rochford, 1992)所谓的“体验的政治(a politics of experience)”中的(p.99)。

斯诺(Snow, 1980)在描述他研究美国佛教运动的田野经历时讨论了三个核心议题:(a)信息充分性,即研究者是否收集到了足够多的资料来回答研究提出的问题;(b)某些外部促发因素,可能是制度因素、人际关系因素、个人内在因素;(c)逼迫民族志者留在田野的一些障碍,如群体对研究者离开的感受和田野关系的整体强度。

在某些情况下,离开田野也受到研究交易的影响。研究者和守门人(gatekeeper)之间的书面或口头协定通常在研究开始时很重要,但之后就在很大程度上被人遗忘了。但是,交易的性质会影响脱离,有助于形成退出策略。

另一个相关的议题涉及研究者身份的错误表述。虽然大多数研究者在其调查中是公开身份的,但在有些情况下研究者会故意误述,或者不能诚实表述,或者不能信守其承诺和誓言。信息提供者期望研究者能够永远信守承诺。研究结束之时承诺失效,研究者返回其日常生活中,信息提供者会感到被背叛和操纵了。研究者准备脱离田野时,需要考虑与其信息提供者未来的关系,特别是如果他们还想回来进一步作研究或澄清时。

加尔迈尔(Gallmeier,1991)描述了他在一组小联盟冰球队员中进行的田野研究【Field Research】,从而说明了通过与信息提供者保持联系可以获得的益处。此类再访(revisiting)使得民族志学者能够获得缺失资料,跟踪仍在场景中的信息提供者的发展轨迹,借此观察到发生的变化。再访也能够使研究者"将发现带回到田野中"(Rochford, 1992, pp.99-100),以便检验其田野报告的效度。有些民族志者也在与试验和信息提供者分享其发现的方式,事实上在最后的分析和评述中还会征询其意见。通过让被访者进行"最后裁定",民族志者在使用一种被称为成员确认【Member Validation】的效度测量。这种方法的长处在于,它让被研究者有机会对所写的涉及他的内容进行回应,也有助于民族志者比较主位(etic)与客位(emic)视角。

——Charles P. Gallmeier

(高勇译校)

参考文献

Gallmeier, C. P. (1991). Leaving, revisiting, and staying in touch: Neglected issues in field research. In W. B. Shaffir & R. A. Stebbins (Eds.), *Fieldwork experience: Qualitative approaches to social research* (pp. 224-231). London: Sage.

Rochford, E. B., Jr. (1992). On the politics of member validation: Taking findings back to Hare Krishna. In G. Miller & J. A. Holstein (Eds.), *Perspectives on social problems* (Vol. 3, pp. 99-116). Greenwich, CT: JAI.

Snow, D. (1980). The disengagement process: A neglected problem in participant observation research. *Qualitative Sociology*, *3*, 100-122.

Leslie 矩阵(Leslie's Matrix)

如果我们知道人口规模与人口结构,以及现在分年龄段的出生率和存活率(或者死亡率),我们可以推断出将来的人口规模与结构吗?莱斯利(Leslie, 1945)创造性地利用矩阵代数【Matrix Algebra】解决了这一问题。

用矩阵符号表示,时点 t+1 时的人口规模 $\mathbf{N}_{t+1}$,是 Leslie 矩阵 $\mathbf{L}$ 与时点 t 时的人口规模 $\mathbf{N}_t$ 的乘积。

$$\mathbf{N}_{t+1} = \mathbf{L}\mathbf{N}_t$$

也可以用向量符号表示为:

$$\begin{bmatrix} N_{1,t+1} \\ N_{2,t+1} \\ \vdots \\ N_{J,t+1} \end{bmatrix} = \begin{bmatrix} F_1 & F_2 & F_3 & \cdots & F_j \\ S_1 & 0 & 0 & \cdots & 0 \\ 0 & S_2 & 0 & \cdots & 0 \\ 0 & 0 & \vdots & & 0 \\ 0 & \cdots & 0 & S_{j-1} & 0 \end{bmatrix} \begin{bmatrix} N_{1,t} \\ N_{2,t} \\ \vdots \\ N_{J,t} \end{bmatrix}$$

其中 $N_{j,t}$ 代表时点 t 时第 j 个年龄组的人口规模,$j = 1,\cdots,J$;$N_{j,t+1}$ 代表时点 t+1 时第 j 个年龄组的人口规模;S_j 代表个体能够活到下一个年龄组的概率,F_j 代表第 j 个年龄组的出生率。

我们假想一个非常简单的例子,总人口被分为三个年龄组,每个年龄组有 10 个人——因此,$\mathbf{N}'_t$ = [10,10,10]。我们进一

步假设三个年龄组的出生率都是 0,两个分年龄组的存活率为 1(最后一个年龄组的存活率被定义为 0,因为没有人能够活过最后一个年龄组的年龄上限)。这时,$\mathbf{N}'_{t+1}$ = [0,10,10],因为除了最年老的年龄组外,每个人都存活到了下一个年龄组。到 $t+3$ 时,此人群将从地球上消失。熟悉矩阵乘法的读者可以很快地看出,结果可以从上述公式中得到。

在更为现实的情景中,Leslie 矩阵的第一行可以是 30%, 40%, 30%的出生率,前两个年龄组的存活率仍然为 1。在这种情况下,我们会有一个稳定的人口规模与人口结构;也就是说,三个年龄组将永远各自保有 10 个人,保持一个静态人口(stationary population)的状态,这是稳定人口模型【Stable Population Model】的一个特例。但是,在现实世界中,人口不会符合如此完美的模型,但是他们的行为可以用 Leslie 矩阵概括。

虽然 Leslie 矩阵最初是设计用于研究人口和动物的规模数量变动的,但是除了对封闭系统的人口规模进行推断外,它在社会科学中还有更广泛的应用。最近的应用包括移民对美国人口增长率的影响的研究(Liao, 2001),以及被称为"城市传奇"的谣言的传播和流行研究(Noymer, 2001)。只要将年龄换为资历,Leslie 矩阵也可用于研究公司或组织规模的增长。

——Tim Futing Liao

(高勇译校)

参考文献

Leslie, P. H. (1945). On the use of matrices in certain population mathematics. *Biometrika*, *33*, 183-212.

Liao, T. F. (2001). How responsive is U. S. population growth to immigration? A situational sensitivity analysis. *Mathematical Population Studies*, *9*, 217-229.

Noymer, A. (2001). The transmission and persistence of "urban legends": Sociological application of age-structured epidemic models. *Journal of Mathematical Sociology*, *25*, 299-323.

分析层面(Level of Analysis)

社会科学分析可以在微观【Micro】层面上研究个体,也可以在宏观【Macro】层面上研究个体的聚合,如城市、县、省或国家成为分析单位【Unit of Analysis】。不存在死板的定义,研究者可以在非常微观和非常宏观之间选择分析层面,如家庭或邻里。

——Tim Futing Liao

(高勇译校)

测量层次(Level of Measurement)

测量层次界定了给变量【Variable】属性【Attribute】的赋值之间的关系。社会科学研究者传统上使用四种层次:定类、定序、定距、定比。界定这四种测量需要的假定【Assumption】依次增加。例如,定类变量【Nominal】只需要离散的类别来代表其属性,定序测量【Ordinal】需要属性是有序的,定距【Interval】测量要求属性之间距离相

等,定比尺度【Ratio】还要求绝对零值有意义。

——Tim Futing Liao

(高勇译校)

显著性水平(Level of Significance)

显著性水平是研究者在对假设【Hypothesis】进行统计显著性检验【Significance Testing】时设定的某一概率【Probability】水平或某一具体取值,通常也被称为检验的阿尔法显著水平【Alpha Significance Level of a Test】。社会科学中的惯例是,研究者要在0.001,0.01,0.05,0.10水平上汇报犯第一类错误【Type Ⅰ Error】的显著性水平,但有时也会使用稍低或稍高的显著性水平。

——Tim Futing Liao

(高勇译校)

Levene 检验(Levene's Test)

这种检验方法是由霍华德·莱文(Levene,1960)提出的,它首先求每个分值与其所在组的均值的绝对差值,然后对这些绝对差值进行单因素方差分析,这些组中的分值是独立不相关的。它是一种 F 检验,其中组间方差的总体估计被用来和组内方差的总体估计进行比较。

$$F=\frac{\text{组间方差估计}}{\text{组内方差估计}}$$

与组内方差估计相比,组间方差估计越大,F 值就越大,它就越可能统计显著。F 值统计显著意味着各组的方差并不相同,或不是方差齐性的。通过转换原始得分,如取平方根或取对数,可以把方差变得更为均等。

方差分析通常被用于确定一个或多个因素(factor)及其交互作用的均值是否显著地不同于随机条件下的期望结果。其假定之一是各组的方差应当相等,以便可以将其汇总起来得到组内方差的单一估计值。但是,如果一个组或多个组的方差比其他组大,较大的方差会增加组内方差,使得这一效果统计显著的可能性变小。

有多种检验可以确定各组是否方差齐性。其中有些检验要求各组中的案例数相同,或者对非正态分布的数据很敏感。Levene 检验是设计用于那些非正态分布并且各组规模不同的数据的。米利肯和约翰逊(Milliken & Johnson,1984, p. 22)推荐对非正态分布数据使用 Levene 检验。Levene 检验目前是统计软件 SPSS 中唯一的方差齐性检验方法。

在因子多于一个的多因素方差分析(factorial analysis of variance)中,Levene 检验将每个单元格视同为单因素方差分析中的一组。例如,一个多因素方差分析中,一个因素有两组(或两个水平),另一个因素有三组(或三个水平),因此就有六个单元格(2×3=6)。这六个单元格的分析方式与单因素方差分析一样。在协方差分析中,绝对差值是用原始分值与用独立因子和协变量预测分值计算得出的。

罗森塔尔和罗斯诺(Rosenthal & Rosnow,1991, p.340)推荐说,在组内分值

相关时,Levene 检验也可以用于检验方差齐性,方法是对每个得分与其组均值的绝对差值进行单因素重复测量方差分析。

——Duncan Cramer

(高勇译校)

参考文献

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin (Ed.), *Contributions to probability and statistics* (pp. 278-292). Stanford, CA: Stanford University Press.

Milliken, G. A., & Johnson, D. A. (1984). *Analysis of messy data: Vol. 1. Designed experiments*. New York: Van Nostrand Reinhold.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research* (2nd ed.). New York: McGraw-Hill.

生命历程研究(Life Course Research)

许多社会科学家的核心关怀就是理解主要的生活事件,如完成教育、结婚、首次生育、就业,特别是理解进入某一事件的转换,以及在人的一生或生命历程中这些事件之间的关系。生命历程研究者通常会采用动态的视角,因为任何生命历程都可能是过去经验和未来期望的结果。生命历程研究使用的方法很多,既有定量的也有定性的,如生活史方法【Life History Method】和事件史分析【Event History Analysis】。

——Tim Futing Liao

(高勇译校)

参考文献

Giele, J. Z., & Elder, G. H. (1998). *Methods of life course research: Qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.

生活史访谈(Life History Interview)

参见生活故事访谈【Life Story Interview】。

生活史方法(Life History Method)

生活故事就是一个人讲述的关于自己经历的故事。阿特金森(Atkinson, 1998)认为,它是"对作为一个整体的某人全部生活经历的相当完整的叙述,同时突出最重要的部分"(p.8)。在社会科学中,生活故事有多种形式:可长可短,可详可略,可实可虚,可浅可深,可现实可浪漫,可现代亦可后现代。它有五花八门的名称:生活故事(life stories),生活史(life histories)、生活叙事(life narratives)、自述故事(self stories)、自我故事(mystories)、自传(autobiographies)、自传/他传(auto/biographies)、口述史(oral histories)、私人证词(personal testaments)、生活文档(life documents)。这些名称各有侧重,但关心的都是对生活的第一人称叙事(对其多样性的讨论,参看 Denzin, 1989, pp.27-49)。

一种简单而有用的入门区分是"长"

生活故事和“短”生活故事。长生活故事(long life stories)也许被视为这种方法的关键——用他自己的语言对其生活进行的完整叙事。这些故事通常是在很长一段时间里收集的,研究者要进行精心的引导,通常还会辅以整理日记、密切观察其生活、访谈其朋友、阅读信件、看照片等。社会学中首次应用生活史的,一般都归于托马斯和兹纳涅茨基(Thomas & Znaniecki, 1918-1920/1958)的五卷本《身处欧美的波兰农民》(*The Polish Peasant in Europe and America*),这本书中认为生活史“是社会学材料的完美类型”,(pp.1822-1833),这一判断非常有名。这项研究中的一卷提供了一位到芝加哥的波兰流亡者 Wladek Wisniewski 的故事,长达 300 页,是在第一次世界大战前 3 个月写的。在其故事中,Wladek 描述了他在波兰乡村 Lubotyn 的早年生活,他出生在这个乡村的一个农村铁匠家中。他谈到了他早年在学校的生活,他进入了面包师行业,他流动到德国寻找工作,最终到达芝加哥,他在芝加哥的困境。20 年后,这一“经典”研究被社会科学界称赞为“前沿社会学研究与理论分析的最完美展现”(参见 Blumer, 1939/1969, p.vi)。从这项经典研究开始,生活史成为芝加哥社会学和波兰社会学研究工作中的重要工具。与此不同,短生活故事(short life stories)花费时间较少,焦点更为集中,通常作为一个系统之一出版,通常都是通过深度访谈来收集它们,辅以开放式问卷,需要精心的追问,通常要花费半小时到三个小时左右。这种故事一般要比长时间故事焦点更为集中。

完整的、主题的、编撰的

奥尔波特(Allport, 1942)进行了另一种区分,他将生活史写作分为三种形式:完整的、主题的、编撰的。完整生活文档(comprehensive life document)旨在掌握一个人生活的全部,完完全全,从出生到现在,理解一个独特的人的成长。主题生活文档(topical life document)并不旨在掌握一个人生活的全部,而是面对一个具体议题。1920 年代后期的《街头混混》(*The Jack Roller*)中对 Stanley 的研究,就是全篇都在关注他生活中的少年犯罪议题(Shaw, 1996)。文档就是用来说明某个具体问题或议题的。编撰的生活文档(edited life document)不会让研究对象用自己的声音来清晰地讲述故事。作者会讲述,会把研究对象的故事编辑到自己的叙事中——这些文档通常更多是用来作为例证的。这一类型的经典是威廉·詹姆斯(James, 1902/1952)的《宗教经验种种》(*The Varieties of Religious Experience*)。在这本长篇大论的研究中,詹姆斯从一系列宗教体验的案例研究中抽取出一些,按照“习俗”“圣洁”“神秘主义”等主题来进行组织。没有一个生活故事占主要地位,而是以摘录形式出现在许多地方。

自然主义的、加以研究的、反身性的

另一种日益盛行的区分方法可称为自然主义的(naturalistic)、加以研究的(researched)、反身性的(reflexive)生活故事,虽然彼此间会有重叠。自然主义的生活故事(naturalistic life stories)就是自然发生的故事,人们只是将其作为日常生活的一部分讲出来,社会科学家对其未加修饰。自然主义的生活故事并不是人为收集的,而是原本就有的。在那些最流行的出版形式中这种故事到处都是:忏悔录、个人见证、自传。经典著作包括卢梭(Rousseau, 1781/1953)的《忏悔录》(*Confessions*),或者奴隶的叙事,如哈丽雅特·雅各布(Jacobs, 1861)的《一个女奴生活中的故事:自述》(*Incidents in the Life of a*

Slave: Written by Herself）。它们都是未经社会科学家修饰的第一人称叙事。

与此不同，加以研究的生活故事（researched life stories）是研究者带着更大的目标（通常是社会科学目标）特别收集而来的。它们并不是在日常生活中自然发生的，而是必须从研究对象那里询问得来。口述史【Oral History】、社会学的生活史【Life History】、文学传记、心理学案例研究——所有这些都产生了原本可能不会存在的生活故事。前面引述的斯坦利和费拉德克的案例研究都是此类典型例子。研究者的角色对此类活动非常重要：没有研究者，就没有生活故事。在研究者的询问中，研究者建构和产生了这些故事。

反身性和返归性的生活故事（reflexive and recursive life stories）与上述两种生活故事类型不同，它对故事的建构和写作意识对比上述两者更为强烈。多数早期生活故事都传达出这样一种意味：他人在讲述一个生活的故事，但是后来的生活故事变得更为自觉，认为讲故事是一种语言游戏，一种言辞活动，一种写作方式，一种社会建构。无论怎样说，讲故事都不只是讲述一个生活的故事。在此，生活被看作更为丰富地"创作"出来或建构出来的：写者也变成了写作的一部分。在这种视角下，生活故事研究通常包括一个故事讲述者和一个访谈者（在这个过程中扮演着引导者的角色），以及收集到的叙事文本：制作者、讲述者，以及文本。制作者和讲述者在特定的社会情境下共同构建了这个故事。故事在不同的情境中可以有不同的讲法。斯坦利（Stanley，1992）甚至发现，许多常见的区分（如传记与自传）都是虚假的："区分虚构作业、传记和自传并不容易"（p.125）。鉴于此，她硬造出了另一个词"自传/他传"（auto/biography）——"这一术语拒绝简单地区分传记与自传，而是承认它们的共生关系"——她把这一术语应用到了她自己的一些研究中（Stanley，1992，p.127）。

虽然生活故事常常被批评为过分主观、过分注重特殊性、不可概推，但是令人瞩目的是，在20世纪，此类研究已越来越流行。

——Ken Plummer

（高勇译校）

* 也可参见自传【Autobiography】、自传式民族志【Autoethnography】、个案研究【Case Study】、人本主义和人本主义研究【Humanism and Humanistic Research】、诠释性传记【Interpretative Biography】、定性研究中的访谈【Interviewing in Qualitative Research】、生活历程研究【Life Course Research】、生活史访谈【Life History Interview】、叙事分析【Narrative Analysis】、叙事访谈【Narrative Interview】、口述史【Oral History】、定性研究【Qualitative Research】、反身性【Reflexivity】。

参考文献

Atkinson, R. (1998). *The life story interview* (Qualitative Research Methods Series, Vol. 44). Thousand Oaks, CA: Sage.

Blumer, H. (1969). *An empirical appraisal of Thomas and Znaniecki (1918-20) The Polish peasant in Europe and America*. New Brunswick, NJ: Transaction Books. (Original work published 1939)

Denzin, N. K. (1989). *Interpretative biography*. London: Sage.

Jacobs, H. A. (1861). *Incidents in the life of a slave girl: Written by herself*. Boston.

James, W. (1952). *The varieties of religious experience: A study in human nature*. London: Longmans, Green & Co. (Original work published 1902)

Plummer, K. (2001). *Documents of life 2: An invitation to a critical humanism*. London: Sage.

Roberts, B. (2002). *Biographical research*. Buckingham, UK: Open University Press.

Rousseau, J.-J. (1953). *The confessions*. Harmondsworth, UK: Penguin. (Original work published 1781.)

Shaw, C. (1966). *The jack roller*. Chicago: University of Chicago Press.

Smith, S., & Watson, J. (2001). *Reading autobiography: A guide for interpreting life narratives*. Minneapolis: University of Minnesota Press.

Stanley, L. (1992). *The auto/biographical I: The theory and practice of feminist auto/biography*. Manchester, UK: Manchester University Press.

Thomas, W. I., & Znaniecki, F. (1918-1920). *The Polish peasant in Europe and America* (5 vols.). New York: Dover.

Thomas, W. I., & Znaniecki, F. (1958). *The Polish peasant in Europe and America* (2 vols.). New York: Dover. (Original five-volume work published in 1918-1920.)

生活故事访谈(Life Story Interview)

生活故事就是一个人尽可能完整和诚实地讲述出自己经历的生活,内容是他所能记起,也愿意他人知晓的,通常会有别人引导他进行访谈【Interview】来进行讲述;它是对一个人生活经历整体的完整叙述,同时特别突出最重要的方面。生活故事是一个人经历的叙事性提炼。时间可以从出生延展到现在,或者在此之前、于此之后。它包括最重要的事件、经历和对生命的感受。

生活故事、生活史【Life History】、口述史【Oral History】都是相同内容的不同说法。它们之间的差异在于强调和侧重面各有不同。生活史和口述史通常关注于一个人生活中的特定侧面,如工作经历,或者在社区生活某个部分中起到的特殊作用等。生活故事访谈主要甚至唯一的关注点则在于人的整体生活经历。生活故事给讲述的生活引入了秩序和意义,对讲述者和倾听者皆是如此。它是一种更好地理解过去和现在的方法,是一种将个人遗产赠予未来的方式。生活故事给予我们一个有利据点,可以透视一个人如何随时间改变而不断体验和主观性地理解他(她)自己的生活。它让我们能够看到和发现那些从孩童到成人将个体生活的各部分串联起来的线条。

历史背景和学科背景

在过去的传统社区中,故事在人们的生活中扮演着重要角色。如今我们讲述的关于自己生活的故事,还是由大量的人类共通、经久不衰的要素引导,如一些原型和主题。这些永恒的体验在不同时间和场景下以不同的形式进行表达,汇聚起来构成了有着诸多重复的一种生活模式,这种模式又成为生活故事的情节基础。我们的生活故事遵循着这样一种模式或者蓝图,以各自不同的方式代表了生活中各种力量之间的一种平衡。只有在这样一种永恒和普遍的背景下,我们才能真正地理解生活故事访谈的重要性和力量所在,才能理解何以它是我们的本质基础。如同古代的神话,我们的生活故事可以发挥使我们与自身、他人、生活奥秘、周遭万物和谐共处的重要功能。

许多学术领域的研究者对他人的生活故事或者生活中的某些方面进行过访谈,其历史要比我们认为的更长。生活故事访谈是从口述史、生活史和其他民族志【Ethnography】和田野方法中演化而来的。作为一种将生活视为整体的定性研究【Qualitative Research】方法,作为一种对个体生活进行深度研究的方式,生活故事访谈是独一无二的。它有着广泛的跨学科应用,可以详尽地理解个体的生活以及个体在社会中如何发挥各种作用。

社会科学家在1930年代和1940年代开始使用一种将生活视为整体的方法:例如

亨利·A.默里使用生活叙述研究个体生活以便理解人格的发展,高尔顿·奥尔波特使用个人档案来研究个体的人格发展。这种方法在埃里克森(Erikson,1975)对路德和甘地的研究中达到了成熟,埃里克森使用生活史来探讨历史性时刻是如何影响个体生活的。

如今,社会科学家对叙事产生了广泛的兴趣,因为它有助于说明身处社会中的个体的生活。生活的叙事研究是一种重要的跨学科方法路数[乔赛尔森和利布利希在编撰的系列书籍(Josselson & Lieblich,1993—1999)可作为代表],其目标是通过深度研究、方法审视和理论探索来深化对个体生活叙事的理解。

作为一种正在多个学科中获得广泛关注和得到广泛使用的研究工具,生活故事有两种主要的研究路数:建构论【Constructionism】和自然论【Naturalism】。有些叙事研究者认为生活故事是访谈者和被访者之间的一种受情境影响的建构性合作。这种路数强调生活故事是在情境中生成的,而不是客观忠实的、受经历指引的叙述。在这种构建性视角下,对生活故事的评价也不是基于它们多么好地符合研究中的生活经历,而是基于不同的人出于不同描述目的是如何使用这种生活叙述的,这里的人并不限于被考察生活的对象本人(Holstein & Gubrium,2000)。

南缅因大学生活研究中心(University of Southern Main Center for the study of Lives)使用的方法是更加自然主义的,发展背景是包括民俗学、咨询心理学、跨文化人类发展等多学科的,从个体生活的客观展示中探讨文化价值和传统在不同的生命周期阶段如何影响发展。这种路数试图从本人的讲述中理解其生活经历以及与他人的关系。如果我们想了解个体的独特视角,最好的办法就是倾听他自己的讲述。这可以使他本人和其他人都将其生活视为一个随时间变化的整体,结果可能是各个部分构成了连续的整体,也可能某些部分看起来与其他部分并不连续,或者两种都可能存在。

生活故事已经在许多学术圈子中赢得了声誉并被接受。心理学家在理解发展和人格中看到了个人叙事的价值。人类学家在对文化相似性和变异性的测量中将生活史或者个体案例研究【Case Study】视为研究的更佳单位。社会学家使用生活故事来理解和定义关系以及群体互动和成员身份。教育学中,生活故事被用作一种新的学习和教学方法。文学学者使用文本自传来探讨设计、风格、内容、文学主题、个体观点(personal truth)。在使用口述史方法的过程中,历史学家发现生活故事材料是丰富地方历史的重要来源。

生活故事访谈的使用

生活故事与生俱来就是跨学科的,它的许多研究应用可以帮助讲述者和学者更好地理解丰富广泛的社会科学议题。生活故事叙事是宝贵的文本,可以借此了解人们的奋斗,了解自我如何随时间变化而发展,以及如何在社会、文化、历史的一席之地中成为有意义的创造者。生活故事是回答"我是谁"这个问题的最有力方法之一。

生活故事有助于研究者更加清醒地意识到人类社会中可能存在的角色和标准的多样性。它们可以在事物的社会秩序中定义个体的位置,可以通过一个场景的道德、伦理、社会背景来解释和确认人的经历。它们可以提供给研究者故事讲述出来却存在在于故事之外的社会现实的信息。

生活故事可以提供线索,让我们理解人们的奋斗与辉煌,人们的根本价值所在,人们的不懈追寻,人们在何处遭遇挫折,人们又在何处重拾信心。生活故事中的宗教、灵性、世界观、信仰、社区构建都被刻画成生活体验。

生活故事访谈的研究应用是无限的。在任何领域,生活故事本身可以作为公开研究中的核心部分,也可以使用其一部分作为说明一些研究目标的资料。生活故事访谈收集的资料往往会比研究者实际使用的更多,这不仅是一种良好的习惯做法,而且也提供了广泛的信息基础来得出结论。生活故事方法可以用于许多学科当中,也可用于许多实质议题中。长久以来人们发展理论所信赖的那些资料是有偏颇的,为了与之平衡和抵消,研究者需要记录更多的关于女性以及形形色色的群体成员的生活故事。女性的声音需要得到更多的机会被人倾听和理解。因为文化调节着我们讲述故事的方式,我们需要倾听那些来自没有被充分代表的群体中个体的生活故事,以便于在文献中建立一种平衡,提升我们所有人的意识水平和知识水平。

生活故事访谈的方法

虽然可以采用一种相当规范的研究方法,并且从生活故事中可以得到很多重要的资料,但是生活故事访谈的操作中总是存在一定的主观性,甚至偶然性。相同的访谈者可能在面对不同的被访者时基于一些变量提出不同的问题,但最后都能得到一份关于被访者的相当完整的生活故事。不同的访谈者可能使用不同的问题,这取决于他们研究项目的特定关注点。生活故事访谈本质上是一种模板,在不同的局面、情况、场景下可以灵活应用。

例如,《生活故事访谈》(*The Life Story Interview*,Atkinson, 1998)一书中建议的200多个问题可以用来获得一个生活故事。这些问题不是说必须全面都要用到,也不是一种一成不变的结构。它们只是建议提出的问题,面对每个被访者时只需要用到最合适的一些问题。有时实际用到的只是少数几个问题,有时则会用到二三十个问题;不管如何,选择的都是最可能的一组问题。最佳的访谈通常是精心选择的开放式问题【Open-Ended Question】的结果,这些问题引导和鼓励被访者给出详尽而深入的回答。

生活故事访谈可以科学地处理(approached),但最好还是作为一种艺术去实施(carried out)。虽然存在可以使用的结构(一组问题,或者一组问题的一部分),但是每个访谈者会以自己的方式来使用它。访谈的实施,无论结构性的或非结构性的,都会因访谈者不同而不同。特定的被访者是另一个重要的因素。有些生活故事的讲述者会为其生活故事的艺术轮廓添加极其个性的意义、回忆、自己的阐释,但另一些讲述者需要更多帮助和引导来填补细节。

生活故事访谈,如生活研究中心实施的那样,包括如下三个步骤:(a)计划(planning)(访谈之前)——准备访谈,包括理解为什么生活故事会有所助益;(b)进行访谈本身(doing the interview itself)(访谈中)——引导人讲述自身故事,并用录音或录像录制下来;(c)对访谈进行转录和阐释(transcribing and interpreting the interview)(访谈后)——删除访谈者的提问和评论,以及其他重复部分(只剩下讲述者自己的话,以便成为用讲述者自己语言表达出来的、流畅的、彼此关联的叙述)。研究者然后可以把转录的生活故事给讲述者本人来审阅和校对,按他意愿进行一些改动。我们最后得到的,就是一个用讲述者自己的语言表达的、流畅的生活故事。需要进行的编辑工作仅仅是删掉重复部分,或者其他完全不相干的信息。

生活故事访谈的长度可长可短。一般而言,对于这里描述的那种生活故事访谈,往往需要对被访者进行至少两到三次访谈,每次访谈会持续一到一个半小时。访谈的长度通常应该能提供绰绰有余的信息,以便很好地

理解被访者的生活或者理解研究问题。

生活故事访谈的特点之一在于,它不是标准化的研究工具。虽然它既是艺术又是科学,而且在访谈前需要大量的计划和准备,但是生活故事访谈是高度情景化和个性化的方法,它收集的是关于人的经历的定性信息。它需要许多访谈者在进行访谈时进行许多即兴的、个人的判断。它的方向可能在瞬间的激发中被意料之外的回答所决定,或者被生活的特殊叙事结构所决定。生活故事访谈中追寻的是展现故事讲述者的独特声音和经历,这种独特声音和经历又可能在某些点上与普遍性的人类经历融会贯通。

伦理和解释问题

有很多重要的概念和解释议题需要考虑。可能最重要的是,因为我们要求的是真人讲真故事,我们在这个过程中试图帮助和协作讲述者,然后把这些故事带给大众,所以我们必须要令人满意地解决如何在最初意图和最终结果之间保持始终如一这样一个重要的伦理问题。进行生活故事访谈的伦理就是要公正、诚实、透明、正直。发展出来的关系是建立在道德基础上的,以确保讲述者的利益、权利和感受得到保护。

同样重要的还有如何对生活故事访谈进行解读。使用生活故事访谈对人的生活进行叙述性调查,其最终目标是对经历进行解释,这是一件复杂的事情,因为阐释和经历是相对而言的。因此主观性处于故事讲述过程的核心。从对每个生活故事本身的审阅过程中,分析范畴伴随着模式和意义的复杂关系自然涌现,而不是如同定量研究一样从开始就被设定好了。基本的解释指导原则是,故事讲述者应当被视为自身生活的专家权威。生活故事访谈是对我们所考虑的生活经历的主观反思,这就要求有一种与此相适应的信度【Reliability】和效度【Validity】标准。

存在多种可以采用的视角,不同访谈者获得的叙述代表了他们所处的不同位置,正如从侧面和正面画出的肖像画都是忠实的画像一样。在个体叙述中,历史真相可能并不是主要议题,真正重要的是生活故事是否被认为是可靠的,这比客观真实还要重要。

但是,内在一致性还是很重要的。叙事中一部分所言不能和另一部分所言自相矛盾。生活中存在各种各样的不一致,人们的反应可能此时这样、彼时那样,但是他们对于自己经历和行为的故事应当与自己保持一致。

不论研究目标是为了一个特定的问题还是为了更多地从个体视角了解人们的生活和社会,生活故事都是理解人们如何看待自己的经历、自己的生活、与他人互动的极佳方法。如果能够牢记某些主要的价值信念,这里建议的生活故事访谈的方法可以避免许多在研究和出版中遇到的典型悖论。帮助人们用自己的语言讲述自己的故事,这既是研究的清晰起点,也是研究的最终结果。

许多生活故事的展示需要对生活故事讲述者赋予自己故事的个体意义给予尊重和敬意。对生活故事的解释和分析以及评论,即使故事是高度个性化和非常主观性的,也应该遵循这一原则,不管是在同一个故事内分析还是多个故事之间比较。在多数情况和多数学科中,理解生活故事的最佳方法就是发现与之相伴、从中而生、有待发掘的意义。

故事讲述是一个个体的意义构建过程。讲故事往往会使得讲述者第一次意识到这些意义,故事文本承载着本身的意义。除此之外,每个学科或者研究课题都会在故事中寻找自己的理论兴趣或者实质兴趣,特别是在某个生活故事或一系列生活故事在理论或主题视角下进行分析时。生活故事访谈

的真正价值在于它既考虑到和阐明了人们共享的一些生活主题、价值、问题、奋斗，也考虑到和阐明了人们之间存在的差异。

——Robert Atkinson

（高勇译校）

参考文献

Atkinson, R. (1998). *The life story interview* (Qualitative Research Methods Series, Vol. 44). Thousand Oaks, CA: Sage.

Erikson, E. H. (1975). *Life history and the historical moment*. New York: Norton.

Holstein, J. A., & Gubrium, J. F. (2000). *The self we live by: Narrative identity in a postmodern world*. New York: Oxford University Press.

Josselson, R., & Lieblich, A. (Eds.). (1993-1999). *The narrative study of lives* (Vols. 1-6). Newbury Park, CA: Sage.

Titon, J. (1980). The life story. *Journal of American Folklore*, *93*(369), 276-292.

生命表（Life Table）

生命表是一种强大而灵活的分析工具，用以描述人口寿命体系的一般参数，如分年龄段的存活概率和平均存活时间。生命表最常用于人口寿命的分析中，但是它也可以用于分析经历某种或某些"退出事件"的概率因持续时间而有所不同的任何总体。例如，它已经被用于学校招生和毕业、劳动力参与、迁移、婚姻状态、健康与残疾的研究中。生命表的主要特点是，不用调整不同人口的结构差异，就可以比较不同人口的死亡参数（或其他退出过程的参数）。

生命表可以有两种不同类型的阐释：一个同期群【Cohort】的经历或者一个静态人口的经历。如果数据是分年龄时期的（period specific），生命表展现的是一个人工合成的（synthetic）同期群随时间变化而在假想中逐渐发生的退出。如果数据遵循的是一个真实的（出生）同期群的经历，生命表被称为"世代生命表"。人口的世代生命表很少见，因为它需要观察特定年龄的死亡率，观察一直要持续到这个出生同期群中最长寿者的生命结束。

如果生命表中的参数被解释为代表一个静态人口的寿命过程，那么出生人数或人口进入数被假定是稳定的，每一个新的同期群都会随时间流逝经历我们观察到的特定年龄的死亡率或退出率。由于出生人数、死亡人数、特定年龄的死亡率都是稳定的，因此这个假想的静态人口的总人数和年龄分布也是稳定的。

基本的生命表函数

生命表包括一系列按年龄组排列的彼此关联的列。

（1）x ——确切年龄 x 。x 的取值范围是从可能的最小年龄 0 到可能的最大年龄 ω 。在节略的生命表中，单年的年龄区间被聚合成为 n 年区间，x 到 $x+n$ 。对于人口来说，通常第一个年龄区间中 n 等于 1，第二个区间中 n 等于 4，除了最后一个区间外的其余区间中 n 等于 5 或 10，最后一个区间是开放的。

（2）${}_na_x$ ——区间存活成数（seperation fraction）。它指在 x 到 $x+n$ 的年龄范围内死亡的人在此年龄区间内存活的时间成数。除了最年幼和最年老的年龄区间，所有的

$_na_x$ 值通常都被设定为 0.50,这个值假定死亡是在区间中均匀分布的。在一岁时发生的死亡,大多集中在这一时间段刚开始的时候,所以最年幼的年龄组的区间存活成数通常小于 0.50。对于较年轻的年龄组的区间存活成数如何设定,有各种各样的建议(参见 Preston, Heuveline, & Guillot, 2001)。

(3) $_nM_x$ ——观察到的年龄在 x 到 $x+n$ 的人口的死亡率。在如下公式中,大写字母 M, D 和 N 分别指年龄正在 x 到 $x+n$ 的人口的死亡率、死亡数目、年中人口总量。如果生命表描述的是全国人口或子人口,死亡数目通常从生命统计数据(vital statistics)中获得,人口总量从普查数据中获得:

$$_nM_x = \frac{_nD_x}{_nN_x}$$

(4) $_nq_x$ ——生命表的同期群中在年龄区间开端(即年龄为 x)时活着但在抵达年龄区间末端时死亡的比例。它是生命表中最关键的一列。除了最年幼和最年老的年龄组, $_nq_x$ 的值通常用如下公式估计:

$$_nq_x = \frac{_nM_x}{\frac{1}{n} + (1 - {_na_x}) \times {_nM_x}}$$

没有人能够活过可能的最大年龄 ω,所以 $_nq_{\omega-n} = 1.0$。由于婴儿的统计漏查是很普遍的,所以 $_1q_0$ 常是从有关婴儿出生和死亡的生命统计数据中获得的(即婴儿死亡率),而不是从婴儿分年龄的死亡数据中获得的。

(5) l_x ——生命表中同期群在年龄为 x 时存活的人数。l_0 的值通常被设定为 100 000。一般而言, $l_{x+n} = l_x - d_x$ 。

(6) $_nd_x$ ——年龄在 x 到 $x+n$ 之间的人发生的死亡数目。

一般而言, $_nd_x = l_x \times {_nq_x}$。由于在最后一个区间开端时存活的人都会在最后一个区间末端时死去, $_nd_{\omega-n} = l_{\omega-n}$。

(7) $_nL_x$——生命表同期群在年龄 x 到 $x+n$ 之间存活的人年数(the number of person-years),或者静态人口中年龄在 x 到 $x+n$之间的存活人数:

$$_nL_x = (n \cdot l_{x+n}) + ({_na_x} \cdot {_nd_x})$$

估计最后一个年龄区间中存活的人数(或者静态人口中年龄大于等于 $\omega - n$ 的人数)会比较麻烦。通常会使用人口中观察到的分年龄的死亡率来估计它:

$$_nL_{\omega-n} = \frac{_nd_{\omega-n}}{_nM_{\omega-n}}$$

(8) T_x ——生命表中同期群在年龄 x 之后存活的人年总数,或者静态人口中年龄大于等于 x 的存活总人数:

$$T_x = \sum_{x}^{\omega-n} {_nL_x}$$

(9) e_x ——年龄为 x 的人的平均剩余寿命:

$$e_x = \frac{\sum_{x}^{\omega-n} {_nL_x}}{l_x} = \frac{T_x}{l_x}$$

表 1 中提供的示例是基于美国 2000 年的总人口数据而得到的,这是一个特定时期生命表(period-specific life table)。生命表参数表明,根据 2000 年观察到的死亡率,一个合成的同期群成员平均会活到 76.8 岁,但是略超过 1/3 的人(35 958/100 000)会度过其 85 岁生日。这个生命表中得到的静态人口总量是恒定的 7 684 601 人,每年都会

有 100 000 人出生和 100 000 人去世,其中 99 393 人的确切年龄为 0~1 岁。

表 1 根据美国 2000 年人口数据做出的生命表

年龄区间	人口中的死亡率	区间存活成数	在区间中死亡的比例	在确切年龄为 x 时存活的人数	在区间中死亡的人数	静态人口		平均剩余寿命的年数
						在区间中存活的人数	年龄大于等于 x 的存活人数	
x 到 $x+n-1$	${}_nM_x$	${}_na_x$	${}_nq_x$	l_x	${}_nd_x$	${}_nL_x$	T_x	e_x
0~1	—[a]	0.125	0.006 9	100 000	690	99 393	7 684 601	76.8
1~4	0.000 328	0.35	0.001 310 9	99 310	130	396 902	7 585 209	76.4
5~14	0.000 186	0.5	0.001 858 3	99 180	184	990 877	7 188 307	72.5
15~24	0.000 815	0.5	0.008 116 9	98 996	804	985 937	6 197 430	62.6
25~34	0.001 080	0.5	0.010 742 0	98 192	1 055	976 646	5 211 493	53.1
35~44	0.001 997	0.5	0.019 772 6	97 137	1 921	961 769	4 234 847	43.6
45~54	0.004 307	0.5	0.042 162 0	95 217	4 015	932 093	3 273 078	34.4
55~64	0.010 054	0.5	0.095 727 8	91 202	8 731	868 367	2 340 986	25.7
65~74	0.024 329	0.5	0.216 904 6	82 471	17 888	735 272	1 472 618	17.9
75~84	0.056 943	0.5	0.443 234 5	64 583	28 625	502 703	737 346	11.4
85+	0.153 244	—[a]	1.000 000 0	35 958	35 958	234 643	234 643	6.5

a.未计算此单元格的值。

生命表的普遍性特点

生命表同期群的初始人数为 l_0。生命表同期群的所有人都会由于死亡而退出生命表,所以死亡总数 ($\sum_x d_x$) 等于生命表同期群(l_0)的初始人数。静态人口的总人数为 T_0。生命表静态人口的粗死亡率(CDR)和粗出生率(CBR)等于 l_0/T_0。生命表人口中年龄为0时的预期寿命(e_0)是粗死亡率的倒数(1/CDR)。关于单递减生命表(其中只有一种退出)、多递减生命表(其中有两种或多种退出)、增减生命表(其中既有进入也有退出)的生命表函数的详细介绍及其阐释与应用(可以参看 Preston et al., 2001; Shryock, Siegel, & Associates, 1973; Smith, 1992)。他们也详尽地介绍了更为复杂的一些人口学方法,如人口的预测及间接估计和静态稳定人口模型的应用,这些方法都是基于这里介绍的基本生命表函数的。

——Gillian Stevens

(高勇译校)

参考文献

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and modeling population processes*. Oxford, UK: Blackwell.

Shryock, H. S., Siegel, J. S., & Associates. (1973). *The methods and materials of demography*. New York: Academic Press.

Smith, D. P. (1992). *Formal demography*. New York: Plenum.

似然比检验(Likelihood Ratio Test)

似然比检验用于决定通过放松某一或某些约束条件,模型的整体拟合是否可以提高,通常也可以用于检验单个系数或系数的线性组合是否等于特定值。它一般用于比较任意两个嵌套【Nested】模型(但是不能用于非嵌套模型)。似然比检验是基于最大似然【Maximum Likelihood】估计原则的,可以检验两个模型对数据的似然差异是系统的还是可以归为抽样误差【Sampliing Error】。如果非约束模型(用最大似然参数估计估出)的似然为 L^*,约束模型的似然为 LR^*,那么似然比统计量的定义为 $LR = -2\ln(LR^*/L^*)$。这可以写作 $2(\ln L^* - \ln LR^*)$。由于去除模型约束总会增加似然,似然比统计量是正的,可以证明它服从自由度为 m 的卡方分布,在此 m 等于约束数目:($LR \sim X_m^2$)。在约束模型和非约束模型同样好的虚无假设下,LR 的期望值为 0。但是考虑到抽样误差,LR 的期望值为约束数目 m。如果模型间存在系统差异,那么 LR 的观察值会大于 m。用经典的假设检验法,可以用 LR 值进行显著性检验。似然比检验渐近地【Asymptotic】等价于 Wald 检验(W)和拉格朗日乘数检验(Lagrange multiplier, LM),后者也被称为劳氏得分检验(Rao's score test)。在小样本中,检验结果会有所不同,但是在线性约束情况下,可以证明 $LM<LR<W$。

——Frederick J. Boehmke

(高勇译校)

参考文献

Kmenta, J. (2000). *Elements of econometrics* (2nd ed.). Ann Arbor: University of Michigan Press.

Maddala, G. G. (1983). *Limited dependent and qualitative variables in econometrics*. Cambridge, UK: Cambridge University Press.

李克特量表(Likert Scale)

李克特量表(累加评定量表)是用来测量人们态度、信念、情绪、感觉、看法、人格特征及其他心理构念(constructs)非常流行的工具。它让人们在一些题项上以不同强度等级的方式表明他们的态度。这类量表【Scale】有三个特征。首先,每个回答者的总得分由多个题项组合而成。其次,每个题项的题干是有关某个对象(人、地方或事情)的一句陈述。这可能涉及完成量表的本人的某个方面,也可能涉及其他人或对象的某个方面。最后,完成量表的人需要从几项强度等级不同的选项中选择一项。选项可

以是单极的从低到高排列,也可以是双极的从极端正向到极端负向排列。这些回答被加以量化(如从 1~6),然后再把不同题项加总,得到一个可以解释的量化得分。

李克特量表的开发一般要包括五个步骤(Spector, 1992)。首先,仔细界定所感兴趣的构念,澄清其本质和边界。其次,决定量表的格式,写出题项。第三,在小样本【Sample】中试验量表,确保题项清晰且易理解。第四,量表在 100~200(或者更多)的人中实施。然后进行题项分析,以帮助选择彼此相关的题项,构成一个具有内在一致性的量表。计算阿尔法系数,它是对于内在一致性信度的测量,阿尔法系数至少应在0.7以上(Nunnally, 1978),虽然 0.8~0.9 更好一些。最后,进行研究来提供量表在构念效度【Construct Validity】方面的资料,以确保将它理解成为步骤 1 中界定的构念的反映是合理的。

李克特量表中的选项可以有不同的类别数量。最通行的形式是同意—不同意(agree-disagree),回答人表明他们对每个题项的赞同程度。通常,有 5~7 个选项——例如,非常不赞同(赞同)[disagree (agree) very much]、有些不赞同(赞同)[disagree(agree) somewhat]和稍微不赞同(赞同)[disagree (agree) slightly]。其他常见的选项还有频度(如从不、很少、有时、经常[e. g., never, seldom, sometimes, often]和评价(如差、一般、好、极好)[e.g., poor, fair, good, outstanding]。

双极量表中选项数应该是奇数、在中间有一个中立选项,还是应该没有中间选项的偶数,对此问题存在争论。支持奇数的人认为不应该强迫态度暧昧者必须选择这一端或那一端。支持偶数的人指出中立选项常常被回答者滥用(例如用来表示这一题项不适用),它鼓励了人们态度不明确。在实践中,选项是奇数还是偶数导致的结果通常差异不大。

——Paul E. Spector

(高勇译校)

参考文献

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Spector, P. E. (1992). *Summated rating scale construction: An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, Series No. 07-082). Newbury Park, CA: Sage.

LIMDEP

LIMDEP 是一种计算机集成程序,用于估计和分析一系列线性和非线性的用于分析有限取值因变量(LIMited DEPendent Variables)和其他形式数据的计量经济学模型和统计模型,如生存数据和样本选择数据。更多信息请参看该软件的网站。

——Tim Futing Liao

(高勇译校)

线性相依(Linear Dependency)

一组向量,如果其中任何一个向量都能被写作其他向量的线性组合,那么被称为是线性相依的。更一般而言,我们说两个变量是线性相依的,如果:

$$\lambda_1 X_1 + \lambda_2 X_2 = 0$$

其中 λ_1 和 λ_2 都是常数【Constant】，λ_1 或 $\lambda_2 \neq 0$。这一定义可以扩展到 k 个变量，如果下式成立，则完全线性关系（perfect linear relationship）存在：

$$\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_k X_k = 0$$

其中的 $\lambda_1 \cdots \lambda_k$ 不同时为零。

举例来说，试考虑以下数据，其中包括三个表示为包括 n 个值的向量的自变量【Independent Variable】：

$$\begin{array}{ccc} X_1 & X_2 & X_3 \end{array}$$
$$\begin{bmatrix} 4 & 16 & 12 \\ 7 & 30 & 22 \\ 12 & 34 & 29 \\ 13 & 22 & 24 \\ 1 & 50 & 26 \end{bmatrix}$$

在此例中，X_3 是 X_1 和 X_2 的完全线性组合$\left(X_{3i} = \frac{1}{2}X_{2i} + X_{1i}\right)$，因此，这个数据集中存在（完全）线性相依。

（自变量中的）线性相依对于回归系数估计来说是很严重的问题。它违背了 $\mathbf{X}$（自变量矩阵【Matrix】）是秩为 k 的 $n \times k$ 矩阵的假定【Assumption】（即 $\mathbf{X}$ 的列是线性独立的，并且至少存在 k 个观察）。此外，它导致了通常被理解为多重共线性【Multicollinearity】的问题。

线性相依对回归系数【Regression Coefficient】估计的影响取决于相依程度。如果自变量中存在完全线性相依，回归系数就不会有唯一解。为了说明这一点，考虑以下有两个自变量的简单回归模型：

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

令 $X_{2i} = 2X_{1i}$（表明两个自变量间存在完全线性相依）。那么，在这种情况下有：

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 2X_{1i} + u_i$$

或者

$$y_i = \beta_0 + (\beta_1 + 2\beta_2) X_{1i} + u_i$$

从上式中可以清楚看到，如果存在完全线性相依，则不可能估计出 X_1 和 X_2 的独立效应（即 β_1 和 β_2）。

用矩阵形式也可以得到类似结果，如果在自变量矩阵（$\mathbf{X}$）中存在完全线性相依，那么（$\mathbf{X}$）和（$\mathbf{X'X}$）都是奇异矩阵，两者皆不可逆。回想一下以矩阵形式表示的回归系数向量为（$\mathbf{X'X}$）$^{-1}\mathbf{X'y}$，所以我们能清楚地看到，如果完全线性相依存在，这些系数是不能被估计出来的。

在社会科学研究中，很少会碰到自变量间存在完全线性相依的情况。但是，很可能会遇到"近完全"线性相依的情况。近完全线性相依会导致高度的多重共线性，而不是完全多重共线性。

如果不完全共线性存在，仍然可以估计回归系数，估计值仍然具有 BLUE 性质（最佳线性无偏估计量），因为不完全共线性并没有违背经典线性回归模型（Classical Linear Regression Model，CLRM）的假定。但是，会存在一些并非无足轻重的后果。系数的方差【Variance】和协方差【Covariance】会比较大。结果，系数的置信区间【Confidence Interval】会更宽（t 统计量【t-Statistics】常常会不显著），导致研究者在虚无假设不正确时也不能够拒斥它（即犯第二类错误【Type Ⅱ Errors】的可能性较高）。最后，系数估计及其标准误【Standard Error】会对数据中的微小变动很敏感。

——Charles H. Ellis

（高勇译校）

参考文献

Green, W. H. (2000). *Econometric analysis* (4th ed.). Upper Saddle River: Prentice Hall.

Gujarati, D. N. (1995). *Basic econometrics* (3rd ed.). New York: McGraw-Hill.

Kennedy, P. (1998). *A guide to econometrics* (4th ed.). Cambridge: MIT Press.

线性回归(Linear Regression)

线性回归指的是表示某一随机变量【Random Variable】(因变量【Depedent Variable】)的条件均值与一个或多个解释变量(自变量【Indepedent Variable】)的相应取值之间的关系【Relationship】的线性函数【Function】。因变量是由解释变量的固定取值的确定性【Determinism】效应和随机扰动项共同决定的随机变量。我们可以将上述关系表示为 $Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + \varepsilon_i$。

我们既可以谈论总体回归函数,也可以谈论样本回归函数。样本回归函数通常用于估计未知的总体回归函数。如果线性回归模型设定正确,那么总体中 ε_i 的期望值是零。给定此期望值,总体回归函数的无偏估计可以从样本中计算出来,只需要解方程 $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \cdots + \hat{\beta}_k X_{ki}$ 即可。这个方程也可以理解为估计一个条件均值,如 $E(Y_i \mid X_i) = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \cdots + \hat{\beta}_k X_{ki}$。两个方程给出的都是在自变量取固定值时随机【Stochastic】因变量的平均值。线性回归分析的目标是发现能够最好地概括 $\hat{Y}_i$ 或 $E(Y_i \mid X_i)$ 的 $\hat{\beta}$ 值。

"线性回归"这一术语并没有说明具体的估计方法。线性回归可以用多种方式进行估计。最常见的方法是普通最小二乘【Ordinary Least Squares, OLS】。这种方法对于扰动项的概率分布并没有作出明确假定。但是,在试图从样本推论总体回归函数时,无偏有效估计要求遵循高斯-马尔科夫假定。在使用 OLS 时,研究者要选择回归参数,以使样本残差平方和最小化。另一种估计线性回归的方法是最大似然估计【Maximum Likelihood Estimation】。这种方法要求的假定除了 OLS 的假定外,还包括扰动项服从正态分布。在使用最大似然法时,分析者要选择回归参数,以使由所有扰动项的联合概率乘积定义的函数最大化。这被称为联合似然函数,还有另一种方法是矩法。这种方法使用类比原则,借此用矩条件来得到估计值。

至于线性的意义,需要澄清的是,我们并不关心变量的线性,而是参数的线性。对线性回归而言,我们要求因变量的预测值是估计参数的线性函数。例如,假定我们估计了一个样本回归为 $\hat{Y}_i = \hat{\alpha} + \hat{\beta}\sqrt{X_i} + \varepsilon_i$,在这个回归中,$\hat{Y}_i$ 和 X_i 的关系并不是线性的,但是 $\hat{Y}_i$ 和 $\sqrt{X_i}$ 的关系是线性的。如果我们将 $\sqrt{X_i}$ 而不是 X_i 视为我们的解释变量,那么就符合线性假定。但是,这不能同样地适用于所有估计值。例如,假定我们想估计如下函数:

$$\hat{Y}_i = \hat{\alpha} + \frac{\sqrt{\hat{\beta}}}{\hat{\alpha}} X_i + \varepsilon_i$$

这就不是一个线性回归了,它需要非线性回归技术。

——B. Dan Wood
Sung Ho Park
（高勇译校）

参考文献

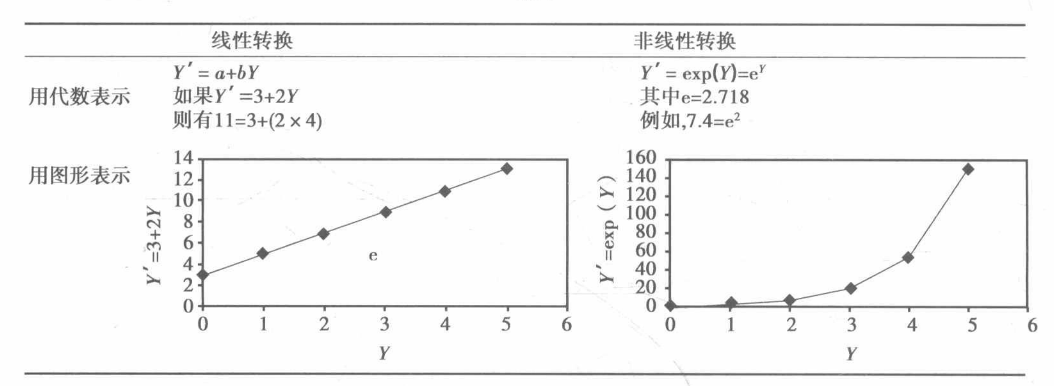
Casella, G., & Berger, R. L. (1990). *Statistical inference*. Belmont, CT: Duxbury.

Gujarati, D. N. (1995). *Basic econometrics* (3rd ed.). New York: McGraw-Hill.

线性转换(Linear Transformation)

线性转换是将一个变量取的所有值转换为一组新的对应值的数学函数。在进行线性转换后，变量的单位和直方图【Histogram】形状都会改变。在统计学中，诸多保秩(rank-preserving)的转换形式被用来从一种分布形态改变到另一种形态。它们又可以分为两类:线性转换和非线性转换,见表 1。非线性转换中的关系是曲线的,而线性转换中的关系是直线的。

表 1



	线性转换	非线性转换
用代数表示	$Y' = a+bY$ 如果$Y' = 3+2Y$ 则有$11=3+(2\times4)$	$Y' = \exp(Y)=e^Y$ 其中e=2.718 例如,$7.4=e^2$
用图形表示		

作为一个函数,如 $X' = 4 + 0.2X$,X 值域上的所有值都应能够唯一地映射到新变量 X' 上,这一点很重要。从华氏到摄氏温度的转换,以及从英制单位到公制单位的转换,都是线性转换。对数转换则是一种非线性转换。

一般而言,变量 X 的线性转换有如下形式:

$$X' = a + bX$$

关于 X 的函数的斜率是一个常数。在实际工作中,常常有必要将收入转换为新的货币单位——如当兑换率是每英镑兑换0.66欧元时,转换函数为 $Y_{euro} = 0.66Y_{£}$,或者当兑换率表示为 R 时转换函数为 $Y_{euro} = RY_{£}$ 。将每周收入转换为年收入也很有用(如 $Y_{yrly} = 52 \cdot Y_{wkly}$)。

更复杂的线性转换会用到不止一个来源变量。包括年度奖金的收入可以用两个变量的内容合并来进行加总:

$$Y_{yrly} = 52 \cdot (H \cdot W) + B$$

其中 H = 每周工作小时数，W = 每小时工资，B = 每年奖金，Y_{yrly} = 总的年度收入。当且仅当 W 对于 H 不变时（即不存在加班补偿时），Y_{yrly} 是 H 的线性转换。总体而言，对于变量 X，转换函数是线性函数的条件为：

$$\frac{D_X^2}{D_X} = 0$$

对不同变量赋予明确权重来构建新量表，这也可以被当作一种线性转换过程。

在生成线性转换时，使用的常数应是非零的。不等于 1 的乘积常数会改变方差【Variance】，变量相对于其均值的相对散布将会以一个比例被改变。最著名的线性转换是一个变量围绕其均值的标准化。要想用与单位无关的语言来表述 X 的变化（也可以说用标准差单位来表述），可以减去 X 的均值再除以其标准差：

$$X'_i = \frac{X_i - \bar{X}}{\mathrm{sd}(X)}$$

其中我们定义

$$\mathrm{sd}(X) = \sqrt{\frac{\sum_i (X_i - \bar{X})^2}{n - 1}}$$

新变量 X' 是 X 的线性转换，因为 X' 对 X 的导数在 X 的所有取值上都是常数。标准差在这里是一个固定参数。对于一个正态分布变量，如教育水平（用年数来近似表达），标准化转换导致 X 分布的变化可以参见图 1 和图 2。

如果 X 的潜在分布是正态的，标准化转换有时也被描述为生成了一个 Z 分布。Z 分布是均值为零、标准差为 1 的标准正态分布。但是，如图 1 和图 2 所示，在这个过程中分布的形态产生了改变。

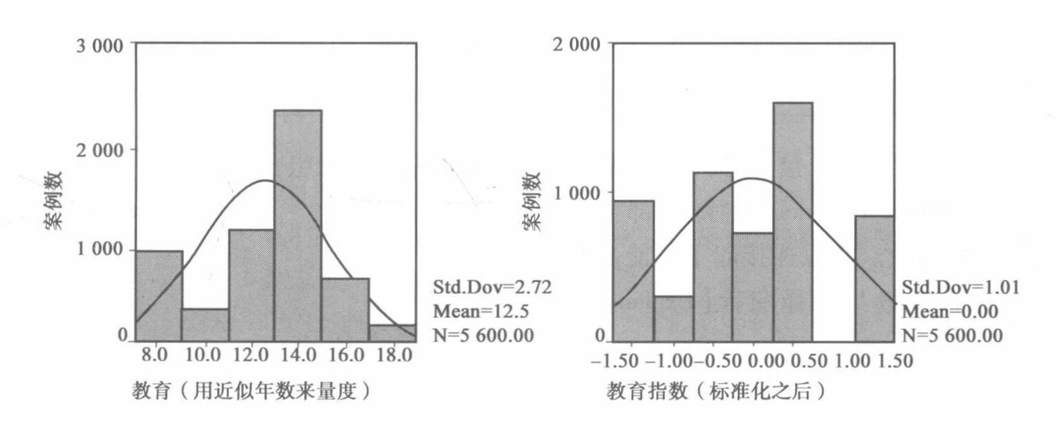


图 1　一个变量的分布及其标准化后的变形

资料来源：英国家户追踪调查，第 9 期，1999—2000（Taylor，2000）。

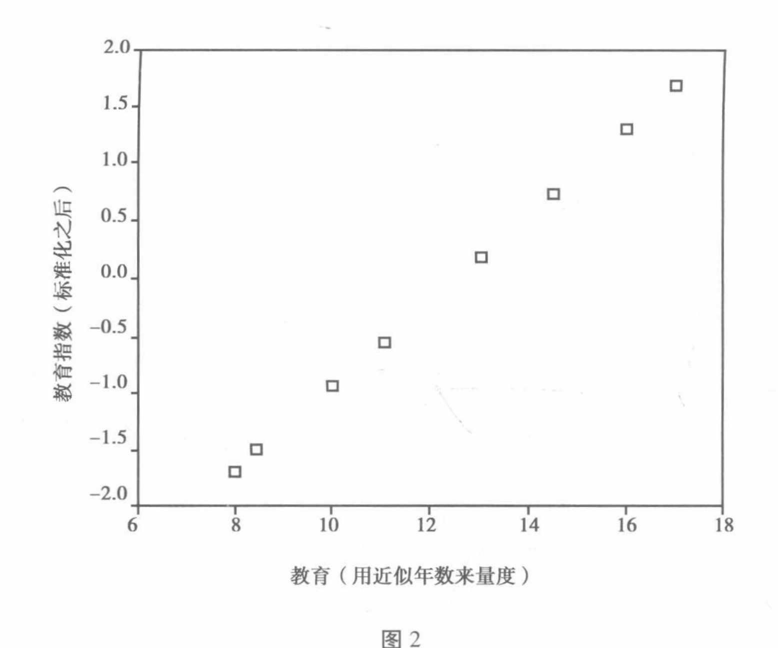


图 2

——Wendy K. Olsen

（高勇译校）

参考文献

Economic and Social Research Council Research Centre on Micro-Social Change. (2001). *British Household Panel Survey* [computer file]. Colchester, UK: The Data Archive [distributor], 28 February. Study Number 4340.

Taylor, M. F. (Ed.) (with Brice, J., Buck, N., & Prentice-Lane, E.). (2000). *British Household Panel Survey User Manual Volume B9: Codebook*. Colchester, UK: University of Essex.

连接函数(Link Function)

连接函数是连接广义线性模型【Generalized Linear Models】的随机部分和系统部分的函数。它也连接起了可能是非线性的结果变量与解释变量【Explanatory Variable】的线性组合。

——Tim Futing Liao

（高勇译校）

LISREL

LISREL 软件是社会科学家广泛用来分析线性结构关系（linear structure relations）模型的首个统计软件包，线性结构关系模型更为人熟知的名称是结构方程模型【Structural Equation Modeling】。如今，它可以分析多种多样的结构方程模型，包

括定序数据和纵贯数据。更多信息，请参看该软件的网页信息。

——Tim Futing Liao

（高勇译校）

整条删除(Listwise Deletion)

整条删除是SPSS等计算机软件程序中与缺失数据【Missing Data】处理相关的一个术语。整条删除缺失数据意味着，分析中涉及的变量存在缺失值的所有个案在分析中均会被忽略。试考虑如下可能情况：一个样本中共有200个个案，我们想得到10个变量的皮尔逊相关系数【Pearson Correlation Coefficient】，但是其中50个个案中至少其中1个变量有缺失值。在这种情况下，就只对剩余的150个个案进行分析。可以与成对删除【Pairwise Deletion】相比较。

——Alan Bryman

（高勇译校）

* 也可参见删除【Deletion】。

文献综述(Literature Review)

文献综述是基于研究证据，对于某些议题或领域所做的总结及/或者与此议题或领域相关的论点总结。有时，综述可以设计成独立成篇，甚至可以成为大部头的书本长度的论著。更常见的是，它们是专著或论文中的章节；在此其功能是为将要报告的专项研究进行铺垫，显示它如何与已知文献相衔接。

最近出现了一些文献综述应当采取什么形式的争论。最常见的是已经被称为“叙述性的(narrative)综述”的形式，这种形式中相关研究的发现和方法被加以概述和讨论，目的在于对从这一领域的知识现状中得到的众多结论提出一个论点。这类综述与“附带注解的参考书目”和所谓“系统性的”综述形成了比照。最近，也有人主张“阐释性的”综述。

一份附带注解的参考书目以简单的方式列出各项研究，可以按字母顺序列出，对每一份研究提供概要，对其优缺点、发现是否有效等进行的评述则可有可无。但是在这种综述和传统的文献综述之间，依据讨论屈从于整体叙述线索的程度，存在着一个连续谱。

最近系统性综述【Systematic Review】很有影响力，显然是因为人们认为它为基于证据的政策制定和实践提供了基础(参见Cooper, 1998; Light & Pillemer, 1984; Slavin, 1986)。系统性的综述有一些与众不同的特点，通常包括如下几点：

- 它们采用明确的程序来决定哪些研究是相关的。
- 它们对研究设计进行明确的层级排列，依据是其产生的研究发现的有效性。基于其是否采用了足够有力的研究设计，或者依据哪项研究提供了最佳的可得证据，一项研究被包括或排除。一般而言，随机对照试验【Randomized Controll Trial】是被最优先考虑的。
- 它们力图在综述中包括所有满足相

关性和有效性规定标准的研究。

• 它们力图综合这些研究中得到的证据,而不是仅仅概述每项研究的发现和以叙述性的或阐述性的方式来讲述这些发现。系统性的综述通常主要关注定量研究,为了综合目的而采用统计学中的元分析【Meta-Analysis】。

在很多方面,系统性的综述是一种将定量方法的工具运用到综述过程本身中的一种尝试(参看 Hammersley, 2001)。

与此相反,阐释性的综述包括了定性的方法路数的应用。在此强调的是综述者的阐释性角色,他在理解不同研究的发现以便对此领域建构一个整体图景,而这一图景正反映了综述者自身特殊的兴趣点和敏感性(参看 Eisenhart, 1998; Schwandt, 1998)。这是从传统的或叙述性的综述中衍生出来的发展。

除此之外,一些评论者主张我们可称之为后现代主义【Postmodernism】的文献综述(参看 Lather, 1999; Livingston, 1999; Meacham, 1998)。在此,任何对现存知识的权威描述都被拒斥,理由是在具体研究或者综述的层次上对真实呈现世界的任何宣称都只不过是将一种散漫的建构强加于其他建构之上;因此反映的是权力的运用或者对权力的抵抗。

不管采用哪种综述方法,下方面的决策必须进行:

• 谁是预定的读者?

• 如何组织综述?

• 怎样寻找相关的研究,哪些研究需要包括进来?

• 对于要讨论的每项研究要提供多少细节;特别是,对于采用的研究方法要给出多少信息?

• 如何评估这些研究及其发现,如何将彼此联系起来?

在许多方面,听众问题是最主要的,因为读者的背景知识和读者的兴趣将影响到对其他问题的回答。最近,很多人非常强调针对研究"使用者"而设计的文献综述——换而言之,强调文献综述是一座桥梁,研究者借此与政策制定者、这样那样的专业实践者、相关大众关联起来。但是,文献综述在研究共同体内部也发挥着重要功能:界定和重新界定学术领域,赋予研究者一种对于参与的工作的集体事业感,使他们与其他人协调其研究工作。

——Martyn Hammersley

(高勇译校)

参考文献

Cooper, H. (1998). *Synthesizing research: A guide for literature reviews* (3rd ed.). Thousand Oaks, CA: Sage.

Eisenhart, M. (1998). On the subject of interpretive reviews. *Review of Educational Research*, *68*(4), 391-399.

Hammersley, M. (2001). On "systematic" reviews of research literatures: A "narrative" reply to Evans and Benefield. *British Educational Research Journal*, *27*(5), 543-554.

Hart, C. (1998). *Doing a literature review*. London: Sage.

Hart, C. (2001). *Doing a literature search*. London: Sage.

Lather, P. (1999). To be of use: The work of reviewing. *Review of Educational Research*, *69*(1), 2-7.

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.

Livingston, G. (1999). Beyond watching over established ways: A review as recasting the literature, recasting the lived. *Review of Educational Research*, *69*(1), 9-19.

Meacham, S. J. (1998). Threads of a new language: A response to Eisenhart's "On the subject of interpretive reviews." *Review of Educational*

Research, *68*(4), 401-407.

Schwandt, T. A. (1998). The interpretive review of educational matters: Is there any other kind? *Review of Educational Research*, *68*(4), 409-412.

Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher*, *15*(9), 5-11.

体验(Lived Experience)

人的经验(experience)是定性研究【Qualitative Research】的主要认识论基础,但是"体验(lived experience)"的概念(从德文Erlebnis翻译而来)在方法论上具有特别的重要性。"体验"一词在狄尔泰(Dilthey,1985)、胡塞尔(Husserl,1970)、梅洛·庞蒂(Merleau-Ponty,1962)等人及其当代倡导者的著作中,表明了一种直接探索人类存在的原初维度或前反思(prereflective)维度的意图。英文中的"experience"一词的词源并不包括"lived"的意思,它源自拉丁文"experientia",意思是"试验、证明、实验、经验"。但是德文中的"Erlebnis"包括了"Leben"(意思是生命或生活)一词。动词"erleben"从字面上说,就是"体验某事"的意思。

狄尔泰(Dilthey,1985)首次对体验及其在人文科学(human sciences)中的重要性进行了系统阐述。他将"体验"描述为反思性或自我给予的(self-given)一种觉悟(awareness),这种觉悟固有地存在于当我们经历其间时生命意识的时间性当中。"唯有在思想中,事物才成为客观"(p. 223),狄尔泰如是说。他认为我们的语言可以被视为一张巨大的言语地图,它为人类体验的可能性命名。

胡塞尔(Husserl,1970)并用"Erlebnis"和"Erfahrung"两词来表示充分成熟的意识行为,在意图性经历(intentional experiences)中意义被赋予。"所有的知识皆始于经验(experience),但并不因此而从经验中诞生",胡塞尔(p. 109)如是说。"Erfahrungen"是一种富于意义的体验,它对我们的存在有着洗心革面式的影响(transformative effect)。更一般而言,经验可以被视为被动地降临、强加、施予我们身上之事,亦可以被理解成更加积极地获取世界某些方面的意义的行动。因此,我们可以谈论一个"有阅历"之人来表明他具有成熟的智慧,而这是生活中不断积累的富于意义和反思的经验的结果。

在《真理与方法》(*Truth and Method*)中,伽达默尔(Gadamer,1975)认为体验有两个意义维度:当下的体验(the immediacy of experience)和已体验之物的内容(the content of what is experienced)。两种意义对于定性探索来说都有方法上的重要性。它们指的是"在所有的解释、加工、沟通之前,当下某事即被掌握"(p. 61)。因此,体验构成了探询、反思、解释的出发点。这一思想在梅洛·庞蒂(Merleau-Ponty,1962)那句名言中也有所反映,"世界并非我之所想,而是我之所活(The world is not what I think, but what I live through)"(pp. xvi-xvii)。要想研究"我之所活"的世界,就必须从"如其所示地直接描述我们的体验"开始(p. vii)。

在当代人文科学中,"体验"仍然是一个核心的方法概念,旨在为人们生活中各种现象的意义提供切实的洞见(参见van Manen, 1997)。例如,医学基于诊断和预判研究模型可以提供干预和行动策略,但是关于疾病的临床发展的预测并不能告诉我们(不同的)人们事实上是如何体验其疾病的。现象学【Phenomenology】的人文科学调

查了人们的体验,来探索疾病的意义的具体维度,如多发性硬化症(Toombs, 2001)或者临床实践当中的疼痛(Madjar, 2001)。

即使如雅克·德里达这样更为语言取向的学者的解构主义【Deconstructionism】和后现代主义【Postmodernism】著作中,体验的概念仍在"经验的独一性"或"绝对存在"这样的用语中得到回响。德里达认为这种原初体验是"对于概念或系统的抵制","这正是我一直准备支持的事情"(Derrida & Ferraris, 2001, p. 40)。在人文科学中,对于经验的关注仍然是突出的,因为它具有突破概括化、编码化、类别计算约束的能力。此外,对于体验的意义与终极来源的关键性探询确保了开放性的存在,而开放性是超越局限来思维和探索的条件。

——Max van Manen

(高勇译校)

参考文献

Derrida, J., & Ferraris, F. (2001). *A taste for the secret*. Cambridge, UK: Polity.

Dilthey, W. (1985). *Poetry and experience: Selected works* (Vol. 5). Princeton, NJ: Princeton University Press.

Gadamer, H.-G. (1975). *Truth and method*. New York: Seabury. Husserl, E. (1970). *Logical investigations* (Vols. 1-2). London: Routledge & Kegan Paul.

Madjar, I. (2001). The lived experience of pain in the context of clinical practice. In S. K. Toombs (Ed.), *Handbook of phenomenology and medicine* (pp. 263-277). Boston: Kluwer Academic.

Merleau-Ponty, M. (1962). *Phenomenology of perception*. London: Routledge & Kegan Paul.

Toombs, S. K. (2001). Reflections on bodily change: The lived experience of disability. In S. K. Toombs (Ed.), *Handbook of phenomenology and medicine* (pp. 247-261). Boston: Kluwer Academic.

van Manen, M. (1997). *Researching lived experience: Human science for an action sensitive pedagogy. London, Ontario: Althouse.*

局部独立(Local Independence)

局部独立是潜变量【Latent Variable】模型的基本假定,包括因子分析【Factor Analysis】、潜特质分析【Latent Trait Model】、潜类分析【Latent Class Analysis】、潜剖面分析【Latent Profile Analysis】等。局部独立假定指的是,给定个体在潜变量上的得分,观察项之间将会彼此独立。这个定义是如下说法的数学表达:潜变量解释了观察项之间彼此相关的原因。

我们用拉扎斯菲尔德和亨利(Lazarsfeld & Henry, 1968)书中的例子来解释局部独立的原则。假定 1 000 个人被问及他们是否读过最近一期的杂志 A 和杂志 B。他们的回答如下:

表 1

	读过杂志 A	没有读过杂志 A	总计
读过杂志 B	260	240	500
没有读过杂志 B	140	360	500
总计	400	600	1 000

很容易证实,两个变量彼此高度相关。杂志 A 的读者更倾向于读杂志 B(65%),而没有读过杂志 A 的人这一比例为 40%。Phi 相关系数,即 2×2 表的积矩相关系数等于 0.245。

假定我们有回答者教育水平的信息，是分为高和低的二分变量。这 1 000 个人被分为这样两组，他们的阅读情况如表 2 所示。在同等的教育水平下，2×2 表中两种杂志之间的相关关系不存在了：每个表的 Phi 系数都等于 0。也就是说，给定教育水平下，杂志 A 和杂志 B 的阅读是彼此独立的。

但是在两个子群体中，阅读行为是非常不同的：高教育组对于两组杂志均有较高的阅读概率（0.60 和 0.80），而低教育组的阅读概率较低（0.20 和 0.20）。A 和 B 之间的相关完全可以用 A 和 B 对于第三个因素（教育）的关系得到解释。注意，A 与 B 列联表的边缘单元格内容，$N(AB)$，可以用如下公式得出：

$$N(AB) = N(H) \cdot P(A \mid H) \cdot P(B \mid H) + N(L) \cdot P(A \mid L) \cdot P(B \mid L)$$

其中 H 和 L 分别表示高教育水平和低教育水平。

表 2②

	高教育水平			低教育水平		
	读过杂志 A	没有读过杂志 A	总计	读过杂志 A	没有读过杂志 A	总计
读过杂志 B	240	160	400	20	80	100
没有读过杂志 B	60	40	100	80	320	400
总计	300	200	500	100	400	500

在潜变量模型中，观察变量被假定在潜变量给定的条件下是局部独立的。这意味着，潜变量的角色和上面例子中的教育水平是相同的。注意，这个假定不仅从实质性角度来讲是有意义的，而且对于确定不可观察的因子来说也是必要的。

局部独立意味着潜变量可以完全解释观察项之间的相关关系，这一事实也给出了对这一假定的简单检验。按照上面给出的 $N(AB)$ 的计算方法，根据模型可以估计出二向表来，它应当和观察到的二向表非常接近。如果指标是连续变量，我们可以比较估计出的相关系数和观察到的相关系数。

已经有各种潜变量模型的变化形式，可以放松一些变量之间的局部独立假定。取决于指标的量度类型，这可以通过在模型中包括直接效应（direct effect）或者彼此相关的误差项（correlated error）来达到。

——Jeroen K. Vermunt

Jay Magidson

（高勇译校）

参考文献

Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis*. London: Arnold.

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.

② 原表 2 中数字有误，无法复合到表 1，译者已经修改。

局部回归(Local Regression)

局部回归是对双变项【Bivariate】或多变项【Multivariate】数据拟合平滑曲线的一种方法。传统方法(如一般最小二乘回归【Regression】分析)是参数法,因为它们需要分析者在拟合曲线前先设定 Y 与 X 之间的函数【Function】形式。与此不同,局部回归是非参数【Nonparametric】法,因为这种算法【Algorithm】试图遵循数据点的经验集聚,无论由此生成的曲线呈现何种形态。平滑曲线的拟合不需要对变量间的函数关系进行任何事先设定。

局部回归的基本原理早在 19 世纪就为人所知。但是它是一种需要大量计算的分析方法,故而直到 1970 年代和 1980 年代早期才得到快速发展。现在最常见的局部回归算法是由威廉克·利快兰及其同事开发的 LOESS 程序。“LOESS”(以前也被称为“LOWESS”)这个词是局部加权回归(locally weighted regression)的首字母缩略。

为数据拟合 LOESS 曲线

假定数据【Data】包括 n 个观察的两个变量 X 和 Y。这些数据可以在一个双变量散点图【Scatterplot】中展示,X 值为水平轴,Y 值为垂直轴。散点图的水平轴上也包括了 m 个点 v_j,其中 j 从 $1 \sim m$。这些 v_j 在 X 的整个值域内均匀分布。

LOESS 用一个“垂直的滑动窗口”沿散点图的水平轴移动。在每一个 v_j 处,这个窗口都会停下来估计一个单独的回归方程(使用加权最小二乘法【Weighted Least Squares】)窗口的宽度被加以调整,以便它总是能够涵盖总数据点中的给定比例(通常称为 α)。在此意义上,m 个回归方程中的每一个都是“局部的”,因为它只包括了落入当前窗口中的 αn 个观察的子集。此外,包括在每个局部回归中的观察都根据其与当前 v_j 的距离进行反比例加权,在当前窗口中靠近 v_j 的观察比远离 v_j 的观察对于局部回归线的位置有着更大的影响。

从每个局部回归中得到的系数被用来对当前窗口估计一个预测值,用 $\hat{g}(v_j)$ 代表。在散点图上可以绘出这 m 个有序配对 $(v_j,\hat{g}(v_j))$,叠加于 n 个数据点上。最后,相邻的拟合点用线段连接起来。这些 v_j 在水平轴上都紧密相接,因此连接起来的线段序列事实上看起来就像是穿过数据点的平滑曲线。观察生成的 LOESS 曲线就可以对 X 和 Y 之间的关系进行实质性的解释。

多变项 LOESS 的基本原则与双变量中的情况完全相同。但是,拟合窗口、v_j 与观察之间的距离、局部加权都是在 k 个自变量生成的 k 维子空间中计算出来的。多变项 LOESS 分析的最终结果是一个平滑曲面,在由自变量和因变量组成的整个 $(k+1)$ 维空间中,这个平滑曲面往往会与数据点云团的中心重合。

LOESS 拟合参数

LOESS 是非参数的拟合程序。但是,分析者在分析前仍然必须要设定一些参数。这些参数值的选择是一个主观过程,但是涉及的考虑因素是很明确的。

第一,平滑参数 α。其定义为每个局部回归中包括的观察比例,它决定了局部拟合窗口的大小。较大的 α 值会生成更为平滑

的 LOESS 曲线。通常的目标是找到能够捕捉数据中主要特征(即穿过数据点中的最密集区域)的 α 值中的最大值。

第二,分析者必须设定局部回归的形式是线性的还是二次方程【Quadratic Equation】的。决定通常要基于对数据的观察。如果数据显示出 X 和 Y 之间的关系总体上是单调【Monotonic】的,那么局部线性拟合就足够了。如果存在非单调的模式,有着局部最小值和/或局部最大值,那么二次局部回归可能会使得拟合曲面更符合数据的波动模式。

第三,分析者必须设定每个局部回归中是否要包括稳健性校正。因为每次局部回归中都只包括了相对较少的数据点,一些异常观察会给系数估计带来不利影响。稳健性步骤会降低任何离群值的影响,以确保拟合出的 LOESS 曲线穿过数据空间中分布最为紧密的区域。

结论

局部回归方法(如 Loess)的最大优点在于它可以非常灵活地发现研究变量之间关系的确切本质。因此,局部回归是探索性数据分析【Exploratory Data Analysis】的重要工具。此外,这种方法的推论【Inference】程序也已经发展出来,使得它在分析数据中可能存在的非线性【Nonlinear】关系特征时非常有用。主要缺点在于,这些方法的转出结果都是图形而非数值,即拟合出的平滑曲线或曲面本身。非参数的拟合方法不能用一个简单的方程式来概括数据。但是,与诸如线性回归分析等传统参数方法相比,局部回归更大程度地让数据"自己说话"。

——William G. Jacoby

(高勇译校)

参考文献

Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, *83*, 596-610.

Fox, J. (2000a). *Multiple and generalized nonparametric regression*. Thousand Oaks, CA: Sage.

Fox, J. (2000b). *Nonparametric simple regression*. Thousand Oaks, CA: Sage.

Jacoby, W. G. (1997). *Statistical graphics for univariate and bivariate data*. Thousand Oaks, CA: Sage.

Loader, C. (1999). *Local regression and likelihood*. New York: Springer.

局部加权回归(Loess)

参见局部回归【Local Regression】。

对数(Logarithm)

对数就是指数。它指的是一个底数——通常是 10 或 e——需要取多少次幂才能得到某一给定数值。因此,如果底数为 10,1 000 的对数就是 3,因为 $10^3 = 1\ 000$。

这一关系也可以用符号表示为 lg 1 000 = 3。对数函数【Function】将变量从原始单位变为对数单位。在上述例子中，1 000 被转换为 3，转换后的单位是以 10 为底的对数单位。最常用的底数有两个，底数 10 和底数 e（近似值为 2.718 2），其中以 e 为底数的自然对数函数在回归模型【Regression Model】中常被用来转换变量。但是，下面的讨论中将既包括常用对数函数（以 10 为底），也包括自然对数函数（以 e 为底），说明的是对数的一般性质。

常用对数函数将变量从 x 个原始单位转换为 y 个以 10 为底的对数单位。这也可以表示为 $y = \lg x$，或者等价地，$10^y = x$。表 1 中展示了以 10 为底的对数函数中，x 值（原始单位）和 y 值（以 10 为底的对数单位）之间的关系。x 的值以乘 10 的速度增加，y 的值以加 1 的速度增加。随着 x 逐渐增大，它要使得 y 值变动一个单位所需的变动幅度越来越大。例如，当 x 从 10 变动为 100 时，y 增加了 1，当 x 从 1 000 变动为 10 000 时，y 仍然只增加了 1。这说明了无论底数为何值，所有对数共有的一个特性：大于 1 的 x 值间的差距在随后的 y 尺度中收缩了。在实践中，这种"收缩"效应可以将较大的正值变小，这样在回归分析中就可以不用删除那些可能干扰分析的游离值【Outlier】。因此，对数函数常用于把有极端正值的变量转换为分布【Distribution】更为对称的变量。

表 1　以 10 为底的对数函数，以及 x 和 y 的关系

x	$y = \lg x$	$10^y = x$	y 的一个单位变动等于 x 的 90%变动
1	0	$10^0 = 1$	
10	1	$10^1 = 10$	{(10−1)/1}×100 = 90%
100	2	$10^2 = 100$	{(100−10)/10}×100 = 90%
1 000	3	$10^3 = 1\ 000$	{(1 000−100)/100}×100 = 90%
10 000	4	$10^4 = 10\ 000$	{(10 000−1 000)/1 000}×100 = 90%
100 000	5	$10^5 = 100\ 000$	{(100 000−10 000)/10 000}×100 = 90%

回归模型中包括对数转换【Transformation】似乎会让回归结果难以解释。但是，由于表 1 最后一列中展示出来的对数特性并不一定如此。对于底数为 10 的对数来说，y 增加 1 个单位（对数单位），x 会稳定地增加 90%（原始单位）。例如，回归模型中以 10 为底的对数被用作解释变量【Explanatory Variable】，相应的回归系数【Regression Coefficient】可以解释为解释变量变动 1 个单位（对数单位）或者变动 90%（原始单位）时结果变量的预期变动。

自然对数函数（以 e 为底数）展示了对数的其他一般特性。对数变换虽然在 x 取大的正值时会缩减差异，但是在 x 取大于 0 小于 1 的值时却会凸现差异。图 1 展示了这一特性，这是一个自然对数函数图，可以表示为 $y = \ln x$，或 $y = \log_e x$。在 x 取值为更小的分数时，随着它越来越逼近于 0，y

以指数速度成为越来越大的负值。这种“延展”特性使得对数函数可以把取值范围为0~1的变量变换为更接近正态分布【Normal Distribution】的变量。图1也说明,如同所有的对数函数一样,自然对数函数在x取负值或零值时是没有定义的。

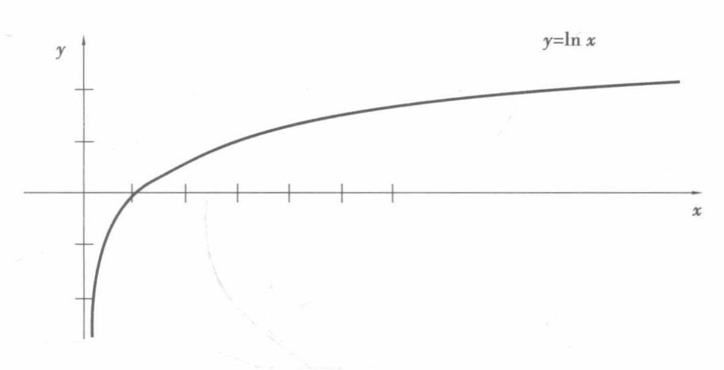


图1 自然对数函数

——Jani S. Little

(高勇译校)

参考文献

Hamilton, L. C. (1990). *Modern data analysis: A first course in applied statistics*. Pacific Grove, CA: Brooks/Cole.

Hamilton, L. C. (1992). *Regression with graphics: A second course in applied statistics*. Pacific Grove, CA: Brooks/Cole.

Pampel, F. C. (2000). *Logistic regression: A primer*. Thousand Oaks, CA: Sage.

Swokowski, E. W. (1979). *Calculus with analytic geometry* (2nd ed.). Boston: Prindle, Weber & Schmidt.

逻辑模型(Logic Model)

逻辑模型是指一些理论假定或模型假定【Assumption】,用来说明项目要素与其预想的中长期后果之间的关系。在研究设计阶段,项目评估者常常用逻辑模型来为资料收集和分析提供整体框架,为评估策略提供概念和指引。模型假定也常被称为“项目理论(program theory)”,描述的是项目内容及原理,项目的工作流程,项目的预期成果。逻辑模型涵盖了项目理论的关键要素,提供了系统的路线图,设定了项目要素的相互关系,说明了可能影响到所预想结果的内部和外部因素。

图1描述了逻辑模型的关键要素,包括如下内容:

- 项目目标。
- 支持项目的可用资源。
- 项目影响的目标总体【Population】或目标人群。
- 构成项目的项目组成部分或关键活动。
- 从项目中得到的项目输出或产出。
- 项目的中间后果和长期后果。
- 项目之外可能影响项目结果的前置变量(antecedent variable)或因素。
- 项目之外可能影响项目设想的短期后果或长期后果的中介事件或干扰事件。

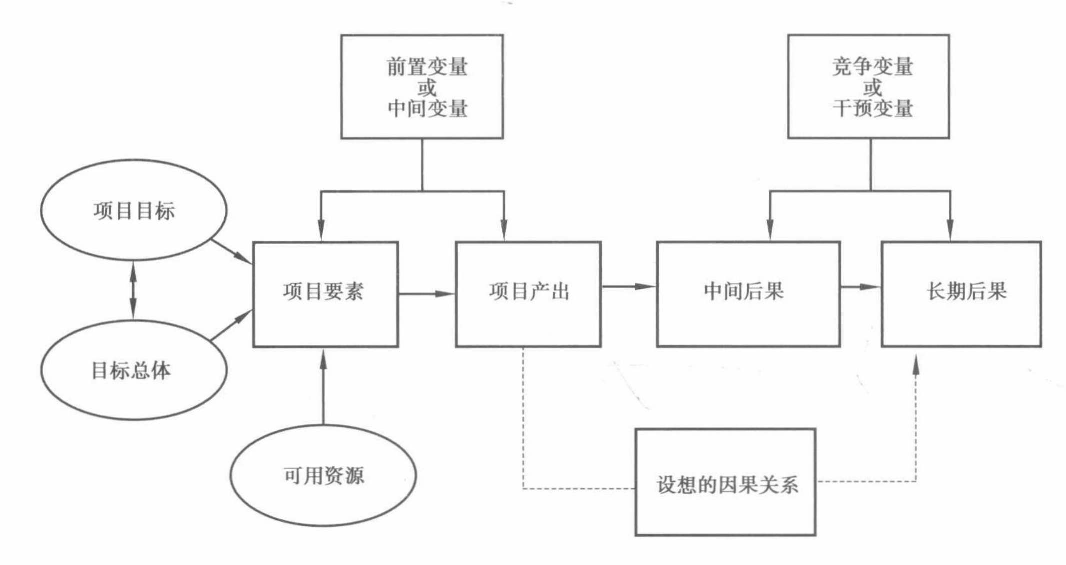


图 1 通用的项目评估逻辑模型

如殷（Yin，1998）所言，对于项目评估者而言，逻辑模型的长处在于，它清晰全面而且足够详尽地描述了事件的链条或彼此关系，有利于对链条中的每一步骤进行操作化测量【Measure】。评估者使用这些操作化测量来制订资料收集策略（即需要收集哪些定量数据和定性资料来全面测量项目的输出或结果）和分析策略（即什么样的中介事件和干扰事件会影响项目结果）。

在全面的逻辑模型中，操作化测量包括从多种来源得到的资料，然后资料进行汇聚，研究者借此更为清晰地了解评估的每一阶段。例如，评估项目输出的中间结果就可以从不同的目标总体或利益相关方进行。与此类似，前置变量或竞争变量（rival variable）可能以不同方式影响到不同的目标人群中的项目输出和结果。评估策略采用多视角或多投入（input）而非单一投入，可以更好地说明项目的预期后果在评估的每一阶段中实现的程度。

除了用于项目评估，项目管理者也在项目设计和项目实施中用逻辑模型来帮助决策、制订策略，报告项目成果时还可以用逻辑模型作为框架。逻辑模型可以用于向利益相关方说明要获得预期成果需要哪些关键因素，哪些事件序列潜在地影响着项目后果，对于不同目标总体何种绩效测量是最相关的，项目在不同目标总体间有何不同后果，有哪些项目不能控制的因素影响了项目参与者或目标总体的中间结果和长期结果，取得预期成果需要哪些资源。

——Georgia T. Karuntzos

（高勇译校）

参考文献

Cooksy，L. J.，Gill，P.，& Kelly，P. A.（2001）. The program logic model as an integrative framework for a multimethod evaluation. *Evaluation and Program Planning*，*24*，119-128.

Julian，D. A.（1997）. The utilization of the logic model as a system level planning and evaluation device. *Evaluation and Program Planning*，*20*(3)，251-257.

Millar，A.，Simeone，R. S.，& Carnevale，J. T.（2001）. Logic models：A system tool for performance management. *Evaluation and Program Planning*，*24*，73-81.

Yin, R. K. (1998). The abridged version of case study research. In L. Bickman & D. J. Rog (Eds.), *Handbook of applied social research methods* (pp. 229-259). Thousand Oaks, CA: Sage.

逻辑经验主义(Logical Empiricism)

参见逻辑实证主义【Logical Positivism】。

逻辑实证主义(Logical Positivism)

逻辑实证主义,有时也被称为逻辑经验主义,是一种生发于1920年代但直到1950年代仍然有极大影响的哲学路数。当时物理学和数学领域的快速发展激励和启迪了逻辑实证主义。逻辑实证主义者的首要任务是分析科学推理和知识的结构。这之所以重要,是因为他们把科学视为智识发展和社会发展的典范。

逻辑实证主义结合了经验主义【Empiricism】和由罗素、弗雷格等创立的新数理逻辑。它的核心命题是,所有有意义的陈述在本质上要么是经验的,要么是逻辑的。这最初被解释为,这些命题要么是有待观察证据检验的,要么是定义问题,因此是同义反复——这一原则被称为可验证性原则。它认为许多哲学讨论和神学讨论完全没有意义,应当排除出去。通过应用这一原则,它希望能够发现真正的问题,而不是那些由于语言误用而产生的问题。常常被排除在外的,不仅包括那些关于上帝或绝对精神(the Absolute)的陈述,而且包括那些经验之外的"真实"世界的陈述。

逻辑实证主义的核心理念之一是对发现情境(the context of discovery)和证明情境(the context of justification)的区分,即对科学家如何产生想法和如何验证科学结论的区分。逻辑实证主义者专注于后者。此外,他们认为科学在方法层面上是一个整体:他们拒绝承认自然科学和社会科学之间存在本质差异。事实上,他们把哲学本身也视为科学的一部分。

逻辑实证主义思想是由维也纳学派和柏林学派提出的,这是由哲学家、数学家、科学家组成的小组,他们在1920年代和1930年代常常聚会。维也纳学派在其中最具影响力,发表了宣言,拥有自己的期刊。尽管主要关注于自然科学,但有几位成员也对社会问题和政治问题很感兴趣。事实上,其中一位成员奥图·纽拉特是社会学家,有一段时间里成了一名社会主义政治家。逻辑实证主义者认为他们的哲学可以在削弱宗教和其他非理性思想来源方面起到重要作用。其他两位重要的20世纪哲学家,维特根斯坦和波普尔,都和维也纳学派有着密切联系——事实上维特根斯坦的《逻辑哲学论》(*Tractatus Logico-Philosophius*)对早期的逻辑实证主义有着很大影响——但是他们两人都是逻辑实证主义的批评者,虽然原因有所不同。在1930年代,由于欧洲大陆上法西斯主义兴起,维也纳学派的主要人物开始移居到英国和美国,这使得逻辑实证主义在英美世界中流行起来,艾耶尔的畅销书《语言、真理与逻辑》(*Language, Truth and Logic*)也对其流行有所作用。

逻辑实证主义从来没有成为统一的思

想体系,它对社会科学方法发展的影响也可能并不像通常认为的那样大;虽然操作化论【Operationism】在20世纪前半叶肯定对心理学有很大影响,对社会学也有所影响。在社会科学中,这些思想与关于科学本质的旧观念有所冲突,特别是关于物理现实的研究和社会现实的研究必须采取不同形式的思想。

逻辑实证主义者自己也意识到了他们最初的主张中存在的问题,并且花了很大力气和心思来努力解决这些问题。一个问题与可验证性原则本身的认识论定位有关:这个原则是有待观察来检验的经验陈述,还是与定义有关的同义反复陈述?它并不能够很容易地划分到某一个类别中,所以有些评论者认为这个原则本身就是自相矛盾的。

另外一个问题源于逻辑实证主义者的经验主义:他们认为存在着非推论性的经验原始材料,如直接观察,可以提供无可置疑的知识基础。但是人们逐渐认识到,任何观察只要被系统地表达出来都会依赖于一些假定,不依赖一些假定这些观察也无法得到检验。因此,实证主义者在理论陈述和经验陈述之间力图作出的鲜明区分并不能成立。

逻辑实证主义者也未能对归纳【Induction】问题提出满意的解答:如何从特殊性的经验材料中得到普遍性的理论结论,同时保有与演绎论证同等的逻辑确定性。事实上,任何经验材料都可以有多种理论解释,从可验证性原则来讲,这可能使得科学理论变得毫无意义。有鉴于此,逻辑实证主义者逐渐用更合理的说法来表达这一原则。结果,他们开始对科学提出了更为现实的叙述,但是这一变化又使得原本清晰的界限变得模糊起来,逻辑实证主义者原本认为那些界限之外的东西都是无意义的形而上学。

直到1950年代,逻辑实证主义都在主导着科学哲学,但自此之后,人们广泛认识到他们面临的问题是无法解决的。因此,一些哲学家(如奎因)抛弃了其核心要素,但保留了其立场中的精神实质。另外一些人(如卡拉托什和贵耶拉本德)采纳了科学哲学的其他解释,如波普尔的解释,努力在此基础上有所建树。也有一些人的反应是采纳了实用主义立场(如罗蒂),或者采纳了维特根斯坦的著作,或者采纳了欧洲大陆哲学运动的立场。对逻辑实证主义的失势也有一些极端的反应,如将科学哲学完全还原为科学史或科学社会学,拒绝将自然科学作为科学的典范,甚至完全放弃了理性的研究方法可以存在这一想法。

——Martyn Hammersley

(高勇译校)

参考文献

Achinstein, P., & Barker, S. F. (1969). *The legacy of logical positivism*. Baltimore: Johns Hopkins University Press.

Ayer, A. J. (1936). *Language, truth and logic*. London: Victor Gollancz.

Friedman, M. (1991). The re-evaluation of logical positivism. *Journal of Philosophy*, *88* (10), 505-519.

Passmore, J. (1967). Logical positivism. In P. Edwards (Ed.), *The encyclopedia of philosophy* (Vol. 5, pp. 52-57). New York: Macmillan.

Reichenbach, H. (1951). *The rise of scientific philosophy*. Berkeley: University of California Press.

Logistic 回归(Logistic Regression)

结果变量(Y)的取值只能是两个离散值——事件发生或不发生,对一道态度问题回答"是"或"否",某一种特征存在或不存在,通常各自编码为1或0(或者1或2)——这种情形在社会科学研究的许多领域都非常普遍,如研究高中生是否会辍学、是否会找到工作、是否会加入工会、是否会生孩子、是否投票给某一位候选人、是否赞成死刑等。Logistic 回归是对这种类型的离散结果进行统计建模的常用工具。

这种技术的整体框架和许多特征基本类似于最小二乘(OLS)回归【Regression】,特别是线性概率模型。OLS 回归中的许多应多加小心之处以及其解决方法(如回归诊断)也同样适用。但是,它的统计假设基础与 OLS 有所不同,更为放宽,以便处理结果变量的有限取值以及引发的问题——因为作为自变量函数的结果变量的取值只有两个;误差【Error】项不再是连续的、同方差的或者正态分布的;预测概率不再是线性约束的,它不可能大于1,也不可能小于0。

Logistic 回归拟合一条特殊的S形曲线,方法是先对解释变量进行线性组合,然后通过 Logistic 函数(参看 Probit 分析【Probit Analysis】)对其进行转换,如下所示:

$$\ln[p/(1-p)] = a + BX + e$$

其中 p 是结果变量取值为1的概率,BX 代表 $b_1x_1+b_2x_2+\cdots+b_nx_n$。模型因此估计的是 Logit,它是结果变量等于1(即事件发生、态度正面、特征存在)或0(即事件未发生、态度负面、特征不存在)的发生比的自然对数。此时,结果变量等于1的概率为:

$$p = [\exp(a+BX)]/[1+\exp(a+BX)]$$

或者

$$p = 1/[1+\exp(-a-BX)]$$

注意,p 和其 Logit 之间的关系是非线性的,如图1所示:在 p 的中段,关系是接近线性的,但是当接近其极值(0或1)时,关系变成了非线性,p 同样程度的变动引起的 Logit 变动越来越大。

对于大样本来说,最大似然估计【Maximum Likelihood Estimation】,MLE 的估计参数是无偏的、有效的、正态分布的,因此可以进行统计显著性检验。这种方法可以推广于结果变量有三个或多个类别(多分类或多项因变量)的情形中。

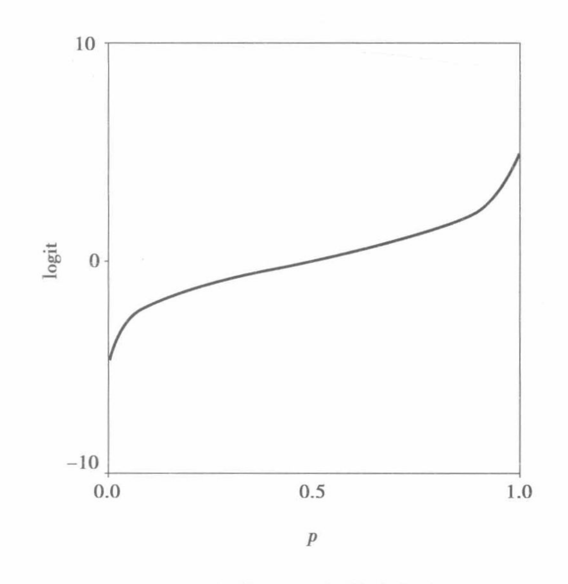


图1 p 和其 Logit 之间的关系

示例

为了说明 Logistic 回归的应用及结果意义,我们使用了综合社会调查(General

Social Survey)中对死刑态度的数据:"你赞成还是反对对犯谋杀罪的人处以死刑?"在1998年的2 599个有效回答中,1 906人(73.3%)表示赞成,693人(26.7%)表示反对。下面的分析考察了种族(白人=1,非白人=0)和教育程度对应因变量【Depedent Variable】的影响,因变量中那些赞成者取值为1,反对者取值为0,控制变量为性别(男性=1,女性=0)和年龄。

标准 Logistic 回归输出包括两个部分:一部分显示模型的总体拟合程度,另一部分报告参数估计。在第一部分中,有几个统计量可以用于比较备择模型,或者评估单个模型的表现。

(1)模型的似然比统计量【Likelihood Ratio】, G^2 或 L^2,用于确定模型相对于只包括常数项的约束模型是否整体上统计显著。它服从 k 个自由度的分布,其中 k 为自变量【Independent Variable】数目。在本例中,我们得到 $G^2 = 156.59$, $df = 4$, $p < 0.000$。模型的卡方表明,预测模型比约束模型有明显提升。

(2)评估 Logistic 回归模型表现的另一种方法需要对观察到的结果和预测出的结果进行交叉分类,然后用正确预测的比例来概括两者之间的对应程度。在本例中,表1显示模型正确地预测了73.3%的案例。总体而言,这个值越大,模型拟合越好。但是这些结果很大程度上取决于结果变量原来的分布。

(3)其他评估模型整体拟合程度的统计量(伪 R^2)类似于 OLS 中的 R^2,但是它们并不能用于假设检验。

表1　交叉分类表

观察到的结果	预测出的结果		正确%
	反对(0)	赞成(1)	
反对(0)	83	605	12.1
赞成(1)	86	1 811	95.5
整体%			73.3

表2　Logistic 回归系数估计

自变量	*B*	*SE*	Wald 统计量	*df*	*p*	Exp
教育程度	−0.056	0.016	11.586	1	0.001	0.946
年龄	−0.006	0.003	4.348	1	0.037	0.994
性别(男=1)	0.481	0.095	25.843	1	0.000	1.617
种族(白种人=1)	1.190	0.107	124.170	1	0.000	3.289
常数	0.922	0.271	11.577	1	0.001	2.514

估计系数通常会在输出中的第二部分给出，格式与 OLS 中类似（参看表 2）。服从自由度为 1 的分布的 Wald 统计量会用于检验系数是否显著地不同于零的假设。在我们的死刑态度例子中，模型中包括的所有自变量在 0.05 水平下都是显著的。但是对系数的解释是完全不同的。Logistic 回归系数表示的是，在控制模型中其他变量的条件下，自变量每一单位变动导致的 Logit 变动。例如，教育的系数为 -0.056，表示受教育年数每增加一年，将会使得 Logit 减少 0.056。另一种更直观的方法是将系数指数化，它表示在优势比上的自变量统计效应。它是乘积性的，因为如果其取值为 1 表明自变量没有效应；取值大于 1 表明存在正效应；取值小于 1 表明存在负效应。例如，在控制其他变量的条件下，受教育年数每增加一年将会使赞成死刑的发生比降低为 0.946 倍。另一种方法是计算一组自变量取值的预测概率。图 2 表明了教育对赞成死刑概率的影响。图中的线分别代表女性非白人（ * ）、男性非白人（▲）、女性白人（+）、男性白人（△），年龄取值固定为样本均值。

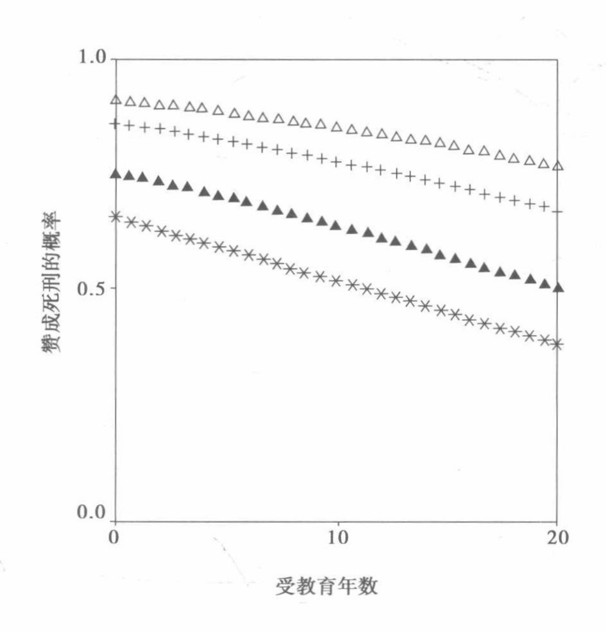


图 2　分教育、种族、性别的赞成死刑概率

——Shin-Kap Han
C. Gray Swicegood
（高勇译校）

参考文献

Liao, T. F. (1994). *Interpreting probability models: Logit, probit, and other generalized linear models*. Thousand Oaks, CA: Sage.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.

Menard, S. (1995). *Applied logistic regression analysis*. Thousand Oaks, CA: Sage.

Pampel, F. C. (2000). *Logistic regression: A primer*. Thousand Oaks, CA: Sage.

Logit

"logit"是事件发生或者状态存在的发生比(odds)的自然对数。例如,如果$P(Y)$是某人使用计算机或者某人犯罪的概率,那么$\text{logit}(Y) = \ln[P(Y)/(1-P(Y))]$。$P(Y)<0.5$时,logit为负;$P(Y)=0$时,logit为负无穷;$P(Y)>0.5$时,logit为正;$P(Y)=1$时,logit为正无穷。$P(Y)=0.5$时,由于1的自然对数为零,$\text{logit}(Y) = \ln[(0.5)/(0.5)] = \ln(1) = 0$。在Logistic回归【Logistic Regression】和Logit模型【Logit Model】中,原来的因变量为二分变量,其logit被用来作为模型中的因变量。

——Scott Menard

(高勇译校)

Logit模型(Logit Model)

logit模型是对数线性模型【Log-Linear Model】的变体,用于分析定类变量,两者之间的关系类似于连续变量【Continuous Variable】分析与普通最小二乘【Ordinary Least Square】回归分析的关系。对数线性模型和logit模型的基本构想始于这样一个假定:我们处理的是一个二维或多维的列联表【Contingency Table】(由两个或多个变量代表),每个维度上都有两个或多个类别,这些类别可以是定序的,但通常被认定是无序的。对分类【Categorical】变量构成的二维列联表的分析是相对简单的问题,基本统计学课本中都会讲到。但是,如果表格包括三个或多个分类变量时,设法在一张纸的二维空间中安排多维表的各种排列方式就会变得困难无比。

对数线性模型用于汇总三维和更高维的列联表,方式是考虑表中每个变量对表中每个单元格中观察频数的影响;它也用于显示表中变量间的关系。如诺克和伯克(Knoke and Burke,1980, p.11)所言,一般对数线性模型(general log-linear model)并不区分自变量和因变量,而是把所有变量都视为自变量,对其彼此关联建模。基于模型的期望单元格频数(如三维表中就是F_{ijk})被认为是模型中所有变量的函数。logit模型在数学上与对数线性模型完全相同,但是其中有一个变量被认为是因变量,而不是对期望单元格频数建模;因变量是某一类别(而非其他类别)的发生比(odds)被建模为其他(自)变量的函数,而不是对期望频数建模。

一般对数线性模型

以一个三维列联表为例,其中变量A为性别(0=男性,1=女性),变量B为种族(1=白人,2=黑人,3=其他),变量C是计算机使用情况,表明被访者是否使用计算机(0=不使用,1=使用)。每个变量的类别标注如下:对变量A而言,$i=0,1$;对变量B而言,$j=1,2,3$;对变量C而言,$k=0,1$。对一个三变量的列联表而言,饱和模型(saturated model)就是包括每个变量的效应,以及模型中变量间所有可能的交互【Interaction】作用效应的模型。使用对数线性模型的符号:

$$F_{ijk} = \eta\,\tau_i^A\,\tau_j^B\,\tau_k^C\,\tau_{ij}^{AB}\,\tau_{jk}^{BC}\,\tau_{ijk}^{ABC} \tag{1}$$

方程左端的每一项都与其他项相乘,各项的意义如下:

F_{ijk}=单元格(i, j, k)的频数,如果模型正确则有望拟合。

η=表中每个单元格频数的几何平均数,$\eta=(f_{010}f_{020}f_{030}f_{110}f_{120}f_{130}f_{011}f_{021}f_{031}f_{111}f_{121}f_{131})^{1/12}$。

τ_i^A=A 对于单元格频数的效应(A 的每个类别都有一个效应,有一个效应是冗余的);类似地,τ_j^B和τ_k^C分别是 B 的效应和 C 的效应。

τ_{ij}^{AB}=A 与 B 的交互效应;A 和 B 不独立而产生的效应;类似地,τ_{ik}^{AC}和τ_{jk}^{BC}分别是 A 与 C 的交互效应和 B 与 C 的交互效应。

τ_{ijk}^{ABC}=A 与 B 与 C 的三维交互效应。

对数线性模型与 logit 模型的记号并不存在单一标准。哥德曼(Goodman,1973)提出了两种记号,第一种记号[方程(1)使用的记号]用系数的指数来表示行和列,这越来越多地用于对数线性分析中。第二种记号用于 logit 模型而非一般的对数线性模型中,表中的行列变量用虚拟变量 $X_1, X_2, \cdots, X_k$ 来表示。第三种记号是拟合边缘记号,常用于对数线性模型而非 logit 模型中,需要用大括号来表明各种效应,例如,将式(1)中的饱和模型表示为 $\{A\}\{B\}\{C\}\{AB\}\{AC\}\{BC\}\{ABC\}$。

在式(1)中的乘积模型中,如果结果中$\tau>1$,则单元格中频数大于几何平均数,如果$\tau<1$,则单元格中频数小于几何平均数。如果$\tau=1$,则此效应对于单元格频数没有影响(单元格频数乘以 1 仍然不变,这类似于 OLS 回归中系数为零)。由于饱和模型包括了所有可能的效应,包括了所有可能的交互作用,它可以完美地重现单元格频数,但是它没有自由度【Degrees of Freedom】。通过约束某些效应参数(τ)等于 1(换言之,从模型中将之排除),我们获得了自由度,因此可以检验模型对观察数据的拟合程度如何。

对式(1)中的饱和模型取自然对数【Logarithm】,可以得到下式:

$$\ln(F_{ijk})=\ln(\eta)+\ln(\tau_i^A)+\ln(\tau_j^B)+\ln(\tau_k^C)+\ln(\tau_{ij}^{AB})+\ln(\tau_{ik}^{AC})+\ln(\tau_{jk}^{BC})+\ln(\tau_{ijk}^{ABC}) \quad (2)$$

或者

$$V_{ijk}=\theta=\lambda_i^A+\lambda_j^B+\lambda_k^C+\lambda_{ij}^{AB}+\lambda_{ik}^{AC}+\lambda_{jk}^{BC}+\lambda_{ijk}^{ABC} \quad (3)$$

其中 $V_{ijk}=\ln(F_{ijk})$,$\theta=\ln(\eta)$ 并且对于相应的每个效应来说 $\lambda=\ln(\tau)$。式(2)和式(3)将式(1)中的乘积模型转换成了与 OLS 回归更类似的加法模型。在式(3)中,如果某个效应的 $\lambda=0$,则表明此效应不存在。例如,如果 A 与 B 之间的交互作用对单元格频数没有影响,则 $\lambda_{ij}^{AB}=0$,这对应于 $\ln(\tau_{ij}^{AB})=0$,因为 $\tau_{ij}^{AB}=1$,则有 $\ln(1)=0$。如诺克和伯克(Knoke & Burke,1980)所言,模型的加法形式有几个优势,包括易于计算标准误差,因此易于决定参数的统计显著性。

Logit 模型

在式(1)、式(2)、式(3)所描述的对数线性模型中,所有的三个变量都有同样的概念地位。对"是否使用计算机"这一变量的处理和性别或种族没有任何差异。但是,从实质意义上讲,我们可能感兴趣的问题是种族或性别对使用计算机与否的影响。Logit 模型是对数线性模型的一种变体,在模型中我们会选择一个变量作为因变量,而将其他变量作为预测变量或自变量。对于式(1)饱和模型中二分因变量 C(是否使用计算机)来说,我们可以将被访者使用计算机的期望比率 Ω_C 表示为使用计算机的期望频率

F_{ij1} 与不使用计算机的期望频率 F_{ij0} 之比：

$$\begin{aligned}\Omega_C &= (F_{ij1}/F_{ij0}) \\ &= (\eta\,\tau_i^A\,\tau_j^B\,\tau_1^C\,\tau_{ij}^{AB}\,\tau_{i1}^{AC}\,\tau_{j1}^{BC}\,\tau_{ij1}^{ABC})/ \\ &\quad (\eta\,\tau_i^A\,\tau_j^B\,\tau_0^C\,\tau_{ij}^{AB}\,\tau_{i0}^{AC}\,\tau_{j0}^{BC}\,\tau_{ij0}^{ABC}) \\ &= (\tau_1^C\,\tau_{i1}^{AC}\,\tau_{j1}^{BC}\,\tau_{ij1}^{ABC})/(\tau_0^C\,\tau_{i0}^{AC}\,\tau_{j0}^{BC}\,\tau_{ij0}^{ABC})\end{aligned} \quad (4)$$

其中的 $\eta, \tau_i^A, \tau_j^B, \tau_{ij}^{AB}$ 从等式的分母和分子中都消掉了，剩下的所有项都包括变量 C。如果我们对式(4)取自然对数，因变量变成了“是否使用计算机”的对数比率，或者被访者是否使用计算机的概率的 logit：$\ln(\Omega_C) = 2\lambda^C + 2\lambda_i^{AC} + 2\lambda_j^{BC} + 2\lambda_{ij}^{ABC}$，或者如果 $\beta = 2\lambda$ 且 $\text{logit}(C) = \ln(\Omega_C)$，则上式可以重新表示为：

$$\begin{aligned}\text{logit}(C) &= \ln(\Omega_C) \\ &= \beta^C + \beta^{AC} + \beta^{BC} + \beta^{ABC}\end{aligned} \quad (5)$$

戈德曼(Goodman, 1973)，哈奇森和索弗罗尼奥(Hutcheson & Sofroniou, 1999)，鲍尔和谢(Powers & Xie, 2000)等人描述或使用过虚拟变量【Dummy Variable】符号来表示上式，形式如下：

$$\begin{aligned}\text{logit}(C) = \beta_0 &+ \beta_1 X_{A=1} + \\ &\beta_2 X_{B=2} + \beta_3 X_{B=3} + \\ &\beta_4 X_{A=1} X_{B=2} + \beta_5 X_{A=1} X_{B=3}\end{aligned} \quad (6)$$

此处，X 变量的下标暗示了哪个类别被作为参照组省略掉了：男性($A=0$)是变量 A 的参照组，白人($B=1$)是变量 B 的参照组。$X_{A=1}$ 是女性的虚拟变量；$X_{B=2}$ 和 $X_{B=3}$ 分别是黑人的虚拟变量和其他种族的虚拟变量；方程中的最后两项表示性别与种族(A 与 B)的交互作用。式(5)和式(6)表示的模型完全相同，它们只是对模型中包括哪些效应的两种不同的表示方式。

在 logit 模型中也可以包括连续、定距或者定比尺度的预测变量(DeMaris, 1992)。有时，人们这样区分 Logistic 回归和 Logit 分析，Logistic 回归是处理连续的预测变量的方法，而 Logit 分析是一种处理类别的预测变量的方法，但这样的区分有些武断。从形式上看，这两种模型是相同的，两者都能够处理连续的和类别的预测变量。它们的区分主要在于估计参数的数值技术不同，它们的表示方式也不同，Logistic 回归模型看起来更像一般最小二乘回归模型，而 Logit 模型看起来更像一般的对数线性模型。

多项变量的 Logit 模型

假如我们现在的因变量不是“是否使用计算机”，而是类别变量“对冰淇淋的偏好”，其中 1=香草味，2=巧克力味，3=草莓味，4=其他。有多种方法可以分析这种类别中不存在自然排序的多项变量，其中一种就是多项 Logit【Multinomial Logit】模型，其中一个类别(如 4=其他)被作为参照组，其他三个类别分别与这个参照组进行比较。由于参照组不能与自身相比，结果会有 $k-1=3$ 个方程，每个方程代表着每个类别与参照组进行相比，我们对式(6)中的系数加入下标来表示我们在拿哪一个类别与参照组相比：

$$\begin{aligned}\text{logit}(C_1) &= \ln[P(C=1)/P(C=4)] \\ &= \ln(F_{ij1}/F_{ij4}) = \beta_1^C + \beta_1^{AC} + \beta_1^{BC} + \beta_1^{ABC} \\ \text{logit}(C_2) &= \ln[P(C=2)/P(C=4)] \\ &= \ln(F_{ij2}/F_{ij4}) = \beta_2^C + \beta_2^{AC} + \beta_2^{BC} + \beta_2^{ABC} \\ \text{logit}(C_3) &= \ln[P(C=3)/P(C=4)] \\ &= \ln(F_{ij3}/F_{ij4}) = \beta_3^C + \beta_3^{AC} + \beta_3^{BC} + \beta_3^{ABC}\end{aligned}$$

其中的 $P(C=1)$ 是预测变量给定时 C 取值为 1 的概率，$P(C=2)$ 是预测变量给定时 C 取值为 2 的概率，以此类推。这些方程表明了性别和种族在多大程度上能够预测人们喜欢香草味 [logit(C_1)]、巧克力味

[logit(C_2)]、草莓味[logit(C_3)]而非其他味道的冰淇淋。多项 logit 模型有($k-1$)组系数,因此解释起来更为复杂,但是它避免了将因变量武断地再编码为一个二分变量的做法。

假定现在因变量 C 代表着不同的学业获得,其中 1 = 没有高中毕业,2 = 高中毕业但没有大学学位;3 = 大学毕业但没有研究生学位;4 = 研究生学位(硕士或博士)。我们可以如前所述构建一个多项模型,但是现在的类别中存在着从低到高的教育水平自然排序,这一信息可以让我们构建一个更简单的模型。如果将第四个类别(研究生学位)作为参照组,我们有理由认为,性别和种族对大学学位与研究生学位之间的差异的影响,与性别和种族对大学学位与高中文凭之间的差异的影响是相同的(也就是说,不论比较的是哪两个相接的教育水平,性别和种族对从一种教育水平升学进入高一级教育水平的影响都是相同的)。如果上述假定成立,我们就可以使用相邻类别 logit(adjacent categories logit)模型,其中 $\ln[P(C=k)/P(C=k+1)] = \sum\beta_k^C + \beta^{AC} + \beta^{BC} + \beta^{ABC}$,或者,对于此处使用的四类别因变量来说,$\text{logit}(C) = \beta_1^C + \beta_2^C + \beta_3^C + \beta^{AC} + \beta^{BC} + \beta^{ABC}$,只有 β^C 的系数有下标,这说明每个类别的 Logit 回归线有着不同的起始点[也被称为界限值(threshold),非常类似于多项 Logit 回归中的截距]。另一种可能是使用累积 Logit(cumulative logit)模型,这一模型的方程与前面模型基本相同,但是因变量使用的 Logit 有所不同:$\ln[P(C\leqslant k)/P(C>k)] = \sum\beta_k^C + \beta^{AC} + \beta^{BC} + \beta^{ABC}$,或者,对于四类别因变量 C = "教育获得"来说,$\text{logit}(C) = \beta_1^C + \beta_2^C + \beta_3^C + \beta^{AC} + \beta^{BC} + \beta^{ABC}$。实际上,相邻类别 Logit 模型和累积 Logit 模型假定每个类别的 Logit 回归线都是平行的,只是起始点有所不同。换言之,这些变量间的关系系数是稳定的,但是在因变量的每个水平上有着不同的起始点。其他一些 Logit 模型也可以用于分析定类因变量,相关评述可以参看 Long(1997)。

(二分的)Logit 模型及其变体 Logistic 回归【Logistic Regression】是最常用于分析二分因变量的两种模型之一,而且比 Probit 模型【Probit Model】更为常用。多项 Logit 模型是如今社会科学家最常用来分析有两个或多个预测变量的多项定类因变量的模型。但是,定序 Logit 模型只是最常用于分析多项定序类别因变量的方法之一。如梅纳德(Menard, 2002)所概括的,其他选择包括定序 Probit 模型,将定序因变量处理为它们代表了定距或定比尺度的变量,定序回归,使用有着多分格相关(polychoric correlation)的结构方程模型。可以对定序 Logit 模型中斜率项是否平行进行检验,如果不平行,使用多项 Logit 模型或者其他模型可能比定序 Logit 模型更恰当,即使数据是定序类别的。广而言之,Logit 模型族,特别是如果我们将 Logistic 回归包括起来,是目前社会科学中分析类别因变量时使用最广泛的方法。

——Scott Menard

(高勇译校)

参考文献

DeMaris, A. (1992). *Logit modeling*. Newbury Park, CA: Sage.

Goodman, L. A. (1973). Causal analysis of data from panel studies and other kinds of surveys. *American Journal of Sociology*, *78*, 1135-1191.

Hutcheson, G., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*. London: Sage.

Knoke, D., & Burke, P. J. (1980). *Log-linear models*. Beverly Hills, CA: Sage.

Long, J. S. (1997). *Regression models for categorical*

and limited dependent variables. Thousand Oaks, CA: Sage.
Menard, S. (2002).*Applied logistic regression analysis* (2nd ed.). Thousand Oaks, CA: Sage.
Powers, D. A., & Xie, Y. (2000).*Statistical methods for categorical data analysis*. San Diego: Academic Press.

对数线性模型(Log-Linear Model)

社会科学中定类变量的交叉表(列联表【Contingency Table】)无处不在,对数线性模型是分析这些表的强大而灵活的工具。对数线性模型可以直接类比于通常用来进行方差分析【Analysis of Variance】的线性模型。当然,有一些重要的问题是列联表分析独有的,这些问题是本词条主要关注的内容。

本词条开始会用 2×2 列联表来展示对数线性模型,从基本概念和这种简单情境中的模型出发,我们会逐步扩展来说明对数线性模型如何用于每个变量的类别超过两个的情况,如何用于变量数超过两个的情况,如何用于两个变量类别间存在一一对应关系的情况。要理解对数线性模型及其应用,优势比【Odds Ratio】的概念和意义是基础,这一点会贯穿始终。

关联:2×2 列联表,对数优势比,对数线性模型参数

相关系数是评估两个连续变量的(线性)关联【Association】的标准测量,经典线性回归【Linear Regression】中的参数与之密切相关。优势比是对数线性模型中的类似测量,对数线性模型还有许多扩展形式,都可以用于定类变量间的关联分析和建模。也就是说,用对数发生比、对数优势比、对数优势比的比较就可以直接解释对数线性模型中的参数。线性模型可类比于方差分析之处在于,模型中的参数可以用均值、均值差、均值比较来进行解释。此外,在测量关联的对数优势比和建立在对数发生比基础上的回归模型参数之间有着密切的联系,正如偏相关【Partial Correlation】系数和普通线性回归中的回归系数【Regression Coefficient】有着密切联系一样。

对数线性模型的基本概念和参数化可以很容易地用2×2 列联表来说明。令 π_{ij} 为在一个 2×2 列联表位于第 i 行第 j 列单元格中的概率,令 n_{ij} 为相应的观察数目。例如,1973 年加州柏克利分校 101 个研究生项目共有 12 763 名申请人,下面是这些申请人的性别和是否被录取的交互表(参见 Agresti, 1984, pp. 71-73):

表 1 柏克利研究生录取数据

性 别	是否被录取	
	是	否
男性	3 738	4 704
女性	1 494	2 827

在此例中,n_{11} = 3 738,n_{12} =4 704, n_{21} = 1 494, n_{22} = 2 827。这些数据被用来分析研究生录取中是否存在性别偏向,因为简单算一下就知道超过 44%的男性被录取(3 738/8 442>0.44),只有不到 35%的女性被录取(1 494/4 321<0.35)。对数线性模型可以用来检验录取决策是(在统计上)独立于申请者性别的,还是两者之间存在关联。

2×2 列联表中单元格概率 π_{ij} 的对数线性模型表达式为：

$$\ln(\pi_{ij}) = \lambda + \lambda_i^R + \lambda_j^C + \lambda_{ij}^{RC}$$

其中的“ln( )”用来表示自然对数函数，λ_i^R，λ_j^C，λ_{ij}^{RC}分别是需要从数据中估计的行参数、列参数和关联参数。统计独立模型就是当 $\lambda_{ij}^{RC}=0$ 对于所有（i，j）都成立时的简化模型。如果 R（行）变量（即申请人性别）与 C（列）变量（即录取决策）并不彼此独立，那么 λ_{ij}^{RC} 就代表了 R 与 C 的关联，这些项目间的比较可以很容易地解释成为评估 R 与 C 的关联程度的对数优势比，即

$$\ln\left(\frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}\right) = \lambda_{11}^{RC} - \lambda_{12}^{RC} - \lambda_{21}^{RC} + \lambda_{22}^{RC}$$

注意，如同方差分析的线性模型中一样，要解出对数线性模型中的个体参数，就必须加入约束条件（例如对于所有 i 和所有 j 来说，$\lambda_1^R=\lambda_1^C=\lambda_{1j}^{RC}=\lambda_{i1}^{RC}=0$，但是无论选择何种对 λ 项的约束条件，$\lambda_{11}^{RC}-\lambda_{12}^{RC}-\lambda_{21}^{RC}+\lambda_{22}^{RC}$ 都是对数优势比。

通常用最大似然法【Maximum Likelihood】来估计对数线性模型及其参数；在柏克利录取数据的例子中，如不假定 R 与 C 是独立的，模型中 $\ln(\pi_{ij})$ 的最大似然估计是 $\ln(n_{ij}/N)$，其中的 N 是总样本量。因此，对数优势比的最大似然估计等于 $\ln\left(\frac{3\ 738 \times 2\ 827}{1\ 494 \times 4\ 704} = 0.41\right)$，检验对数优势比等于零的虚无假设【Null Hypothesis】的统计量（计算细节请参看优势比【Odds Ratio】）等于 10.52，因此这提供了强有力的证据（$p<0.001$）表明两者并非独立的，而是存在相关的。

在对数线性模型框架中进行假设检验的一般方法是，在虚无假设（上例中为两者独立）和备择假设【Alternative Hypothesis】（上例中为包括了 λ_{ij}^{RC} 项的非约束模型）条件下拟合模型。两个模型的似然比统计量【Likelihood Ratio】（用 L^2 表示）的差值就是对虚无假设和备择假设何者更可信的检验，（在特定假定下）合适的参照分布为卡方分布，自由度【Degrees of Freedom】（在此表示为 df）等于两个模型的（残差）自由度差值。在本例中，虚无假设有 $L^2=112.40$，$df=1$。本例中的非约束模型中没有进行简化假定，因此残差偏差和残差自由度都等于 0。因此，用于检验独立虚无假设和非约束性备择假设的 L^2 等于 112.40，相应的 p 值远小于 0.001。

注意，L^2 的平方根（即 $\sqrt{112.40}=10.60$）与对数优势比的最大似然估计的检验统计量在数值上非常接近。事实上，虽然前者基于数据估计及其估计标准误差，后者基于对数似然，但两种检验统计量在数值上的接近并不是偶然的。这两种对对数线性模型进行假设检验的方法是渐进等效的，直观而言，这意味着如果观察的样本量很“大”时它们的数值会非常接近。

对数线性模型是所谓广义线性模型【Generalized Linear Model】族中的一种，因此适用于此类模型的软件（如 R 或 Splus 中的 glm，或者 SAS 中的 proc glm）都可以用来拟合和分析对数线性模型。

如果视观察到的交叉分类的抽样模型为多项分布【Multinomial Distribution】、乘积多项分布或者独立泊松分布【Poisson】，那么就可以将基于似然推断的标准理论用于对数线性模型估计和检验。使用上述三种抽样模型中的哪一种，对推理结果影响不大，只要估计出来的对数线性模型不违反抽样假定。例如，在乘积多项抽样的情形中，估计出的模型得到的单元格频数估计必须保持相关的多项总数。对这些问题的详尽

讨论,可以参看本词条中有关一般类别数据分析的参考文献。

多于两个变量的模型

对数线性模型在类别变量多于两个情形中的扩展,既包括方程中的模型设定,也包括对独立的相关概念的发展。在此的分析仍然会进一步关注加州大学柏克利分校的研究生录取数据。下面的数据中包括了第三个变量(除了申请人的性别和录取状态外):申请的专业。展示的数据只包括了当年六个最大的研究生专业,包括了 4 526 名申请人,占全体申请人的 35%。

包括这三个变量的对数线性模型的最完全设定为:

$$\ln(\pi_{ijk}) = \lambda + \lambda_i^R + \lambda_j^C + \lambda_k^M + \lambda_{ij}^{RC} + \lambda_{ik}^{RM} + \lambda_{jk}^{RM} + \lambda_{ijk}^{RCM}$$

表 2　六个专业的柏克利录取数据

专　业	是否被录取			
	男　性		女　性	
	是	否	是	否
A	512	313	89	19
B	353	207	17	8
C	120	205	202	391
D	138	279	131	244
E	53	138	94	299
F	22	351	24	317

其中的下标 k 表示第三个变量。沿用前面例子中的符号(即 R 代表申请人的性别,C 代表是否被录取),M 用来代表申请的专业。用来探索某一特定专业中 R 与 C 的关联的对数优势比为:

$$\ln\left(\frac{\pi_{11k}\pi_{22k}}{\pi_{12k}\pi_{21k}}\right) = \ln(\pi_{11k}) - \ln(\pi_{12k}) - \ln(\pi_{21k}) + \ln(\pi_{22k})$$

上式又可以写为:

$$\ln\left(\frac{\pi_{11k}\pi_{22k}}{\pi_{12k}\pi_{21k}}\right) = (\lambda_{11}^{RC} - \lambda_{12}^{RC} - \lambda_{21}^{RC} + \lambda_{22}^{RC}) + (\lambda_{11k}^{RCM} - \lambda_{12k}^{RCM} - \lambda_{21k}^{RCM} + \lambda_{22k}^{RCM})$$

要注意,在后一个等式中的第一个式子并不涉及 M 的类别,第二个式子则会涉及。令 $\theta_k^{RC(M)}$ 代表第 k 个专业中性别与录取与否之间的条件关联的对数优势比[即 $\theta_k^{RC(M)} = \ln(\pi_{11k}\pi_{22k}/\pi_{12k}\pi_{21k})$],令

$$\delta^{RC} = \lambda_{11}^{RC} - \lambda_{12}^{RC} - \lambda_{21}^{RC} + \lambda_{22}^{RC}$$

且

$$\delta_k^{RCM} = \lambda_{11k}^{RCM} - \lambda_{12k}^{RCM} - \lambda_{21k}^{RCM} + \lambda_{22k}^{RCM}$$

因此有 $\theta_k^{RC(M)} = \delta^{RC} + \delta_k^{RCM}$。显而易见，这表明全模型允许 R 与 C 的关联可以在 M 的不同类别中有所不同；也就是说，只要 δ_k^{RCM} 不为零，$\theta_k^{RC(M)}$ 在不同专业中就是不同的。另一方面，去掉了三因素项 λ_{ijk}^{RCM} 的模型（即 δ_k^{RCM} 为零）就假定 R 与 C 的关联在不同专业中都是相同的——这是同质关联模型（model of homogeneous association），或者无三因素交互模型（model of no-three-factor interaction）。类似，去掉了 λ_{ij}^{RC} 和 λ_{ijk}^{RCM} 的模型产生的 R-C 关联的对数优势比都是 0，对 $M = k$ 都是如此，因此这是一个假定 R 和 C 在六个专业中都是独立的模型。后一个模型就是条件独立（conditional independence）模型，R 和 C 在每一个水平 $M=k$ 中都是独立的。

同质关联模型的似然比检验统计量为 $L^2 = 20.20$，残差 df 为 5（即六个对数优势比被约束为一个值，因此 $df = 6 - 1$）。相应的 p 值接近于 0.001。表明数据是有些背离同质关联模型的。对数据拟合异质关联模型（即全模型）得到的是数据的完美拟合，对于估计值 $\hat{\delta}_k^{RCM}$ 的考察（此处未展示）表明专业 A 中的 R 与 C 的关联与其他五个专业有着显著不同。事实上，没有证据能够表明，其他专业中 R 与 C 存在关联。因此，考虑拟合如下模型，其中 δ_k^{RCM} 只对 $k = 1$ 进行估计；也就是说，对于所有不等于 1 的 k 来说，我们约束 $\delta_k^{RCM} = 0$。似然比模型统计量为 $L^2 = 2.68$，残差 df 为 5（我们在六个可能的对数优势比中只估计了一个），这表明对数据的拟合非常好。

在此没有篇幅来对柏克利录取数据进行更为详尽的分析，但是值得注意的是，一前一后两个例子是辛普森悖论（Simpson's paradox）很好的说明。也就是说，在不考虑专业时，申请人的性别和录取与否是有关联的，这会让我们倾向于认为存在对女性申请人的歧视，但是如果我们控制了专业，就会发现不存在对某一个性别的整体偏向。数据表明，在接收申请人最多的那个系中，女性是以更高比率被录取的，这与不分专业的两向表中观察到的偏向方向是相反的。悖论之所以存在，是因为男性与女性申请各个专业的比例并不相同。也就是说，女性更可能申请那些低录取率的专业，因此如果不分专业来看时她们的总录取率就显得更低。

多于两个类别的对数线性模型：两向示例与 λ_{ij}^{RC} 参数的精练与修正

在 $I \times J$ 交叉分类的对数线性模型的一般设定中，参数仍然可以用关联的对数优势比来解释；更一般地，在多向表中，它们可以用条件对数优势比或偏对数优势比来进行解释。例如，在一个任意的两向表中，关联参数 λ_{ij}^{RC} 的对比就对应于一个对数优势比，这个优势比比较某一个变量的两个类别在另一个变量的两个类别上的相对比率，即

$$\ln\left(\frac{\pi_{ij}\pi_{i'j'}}{\pi_{ij'}\pi_{i'j}}\right) = \lambda_{ij}^{RC} - \lambda_{ij'}^{RC} - \lambda_{i'j}^{RC} + \lambda_{jj'}^{RC}$$

考虑如下从教育获得研究中得到的数据。数据来自 1973 年第二世代职业变迁调查（Occupational Changes in a Generational Ⅱ Survey）（Mare, 1994），在此我们只考虑父亲至少接受过 16 年教育的那些子代。行对应于调查被访者的教育获得，列对应于被访者的兄长的教育获得。检验被访者的教育获得与其兄长的教育获得之间统计独立的似然比统计量为 $L^2 = 161.86$。自由度为 4 [即$(3 - 1) \times (3 - 1) = 4$]，在相应的自由度为 4 中的卡方分布中，$p$ 值远小于0.001，因此有强证据表明两者并非独立的。我们

可以拟合饱和模型来估计四个可识别的 λ_{ij}^{RC}，但是我们在此使用这些数据来说明，如何通过对 λ_{ij}^{RC} 施加约束条件来得到更为简约的模型，以便探索数据在哪些具体方面是偏离独立的。

表 3 教育获得数据

被访者的受教育年数	兄长的受教育年数		
	<12	12	≥13
<12	16	20	16
12	15	65	76
≥13	26	69	629

我们考虑的第一个模型利用了两个变量的类别都是定序的这一事实。令 $\lambda_{ij}^{RC} = \theta_{ij}$，因此比较任何一对相邻类别的对数优势比都等于 θ，即

$$\ln\left(\frac{\pi_{ij}\pi_{i+1,j+1}}{\pi_{i,j+1}\pi_{i+1,j}}\right) = \theta$$

此类模型被称为关联模型【Association Models】的一个例子，在本例中，这个模型的 $L^2 = 33.12$，自由度为 3。这在拟合上是很大的进步，但是仍然不太理想。

下一步，考虑进一步精练模型，在模型中加入一个参数，它只对在主对角线上的观察有效，即 $\lambda_{ij}^{RC} = \theta_{ij} + \beta I(i=j)$，其中的 $I(i=j)$ 在 $i=j$ 时取值为 1，否则取值为零。这个模型进一步提升了拟合（$L^2 = 6.36$，自由度为 2，相应的 p 值为 0.025～0.05），估计出来的关联参数为 $\hat{\theta} = 0.47$ 和 $\hat{\beta} = 0.74$。简言之，$\hat{\theta}$ 说明被访者的教育获得与其兄长的教育获得之间存在正向关联（即 $\theta > 0$），$\hat{\beta}$ 说明两个变量更多地聚集在同一类别（即对角线上）而非呈现均匀的关联关系（即 $\beta > 0$）。

在此提出的两个精练模型仅仅初步触及了对数线性模型框架内对关联关系建模的工具。事实上，对称【Symmetry】模型（即 $\pi_{ij} = \pi_{ji}$）对数据的拟合很好（即 $L^2 = 3.46$，自由度为 3；$p > 0.25$），这种模型对应于 $\lambda_i^R = \lambda_i^C$ 且 $\lambda_{ij}^{RC} = \lambda_{ji}^{RC}$ 时的对数线性模型。对此感兴趣的读者可以看参考文献以及这些参考文献中引用的文献来进一步探索这一问题，包括两向表和多向表中的关联关系如何扩展到非线性模型中。

——Mark P. Becker

（高勇译校）

参考文献

Agresti, A. (1984). *Analysis of ordinal categorical data*. New York: John Wiley.

Agresti, A. (1996). *An introduction to categorical data analysis*. New York: John Wiley.

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: John Wiley.

Mare, R. D. (1994). Discrete-time bivariate hazards with unobserved heterogeneity: A partially observed contingency table approach. In P. V. Marsden (Ed.), *Sociological methodology 1984* (pp. 341-383). Cambridge, MA: Blackwell.

Powers, D. A., & Xie, Y. (2000). *Statistical methods for categorical data analysis*. San Diego: Academic Press.

纵贯研究(Longitudinal Research)

纵贯研究最好与横截面【Cross-Sectional】研究进行比较来理解。横截面研究指收集一组案例(可以是个体或个体集合,如城市或国家)在一些变量(如非法行为频数、对全球化的态度等)上的数据的研究,其特点在于(a)数据收集发生在一个时间点上,(b)收集的数据只是某一个时间点或某一个时间段的信息(因此都是某一"时期"的信息)。纯粹的横截面数据可以用来考察案例间差异,但是不能考察案例的变化。不同学科对纵贯研究有不同定义,但是宽泛的定义包括有如下特征的研究:(a)收集的数据是多个时期的信息;(b)可能在不同时期进行数据收集工作(并非必要条件);(c)可以对同一案例(个体或个体集合)随时间发生的变化进行测量和分析。国家层面的纵贯数据的收集始于1665年,已经有300年历史了,个体层面的收集始于1759年,但是社会科学中纵贯研究的兴盛始于1970年代和1980年代(Menard, 2002)。

有些研究者认为在不同时间点收集数据是纵贯研究的必要特征,但是原则上来讲纵贯研究可以收集回溯性(retrospective)数据,即在某一个时期收集关于多个时期的数据。例如,被访者可以讲述自己被罪犯侵害的所有经历,或者自己犯罪的所有经历,时间跨度可以是很多年甚至其一生。与回溯性研究不同,前瞻性(prospective)纵贯数据不断进行数据收集,被访者只需回忆最近时期发生的事情:上周或上月或上年的被侵害经历或犯罪经历,或者数据收集时的态度、信仰、社会人口特征。

纵贯研究有两个主要目标:描述变迁模式,确立因果关系的方向(正或负,*Y*影响*X*或*X*影响*Y*)与强度(强度为零表明不存在因果关系)。变迁通常是从如下两种连续体之一的角度进行测量的:编年性的时间(如历史变迁)或者年龄时间(如成长变迁)。有时,很难把两者彻底分开。如果老人比年轻人的犯罪行为更少,这是由于犯罪行为随年龄而减少呢,还是由于那些老人一直很少犯罪(甚至他们年轻时也犯罪很少),而那些年轻人仍会更多犯罪(甚至他们变老时也会犯罪较多)?用横截面数据不可能区分历史效应和成长效应。即使用纵贯数据也很难分清两者,但是如果有对相同个体在不同年龄和不同历史时期的数据,就至少有可能分清这两种效应。

纵贯设计可以分为几种不同类型(Menard, 2002),如图1所示。在图1的每一部分中,列代表年,行代表研究对象,用进入研究的不同时间来分组。因此,研究对象在不同时间点(列)进入总体或样本(行)中,那些在不同时间点进入研究的研究对象可能同时处在研究中(同一列中可以有多行),但是重复性横截面设计(repeated cross-sectional design)中不同时间点上研究的对象是不同的(同一列中只有一行)。

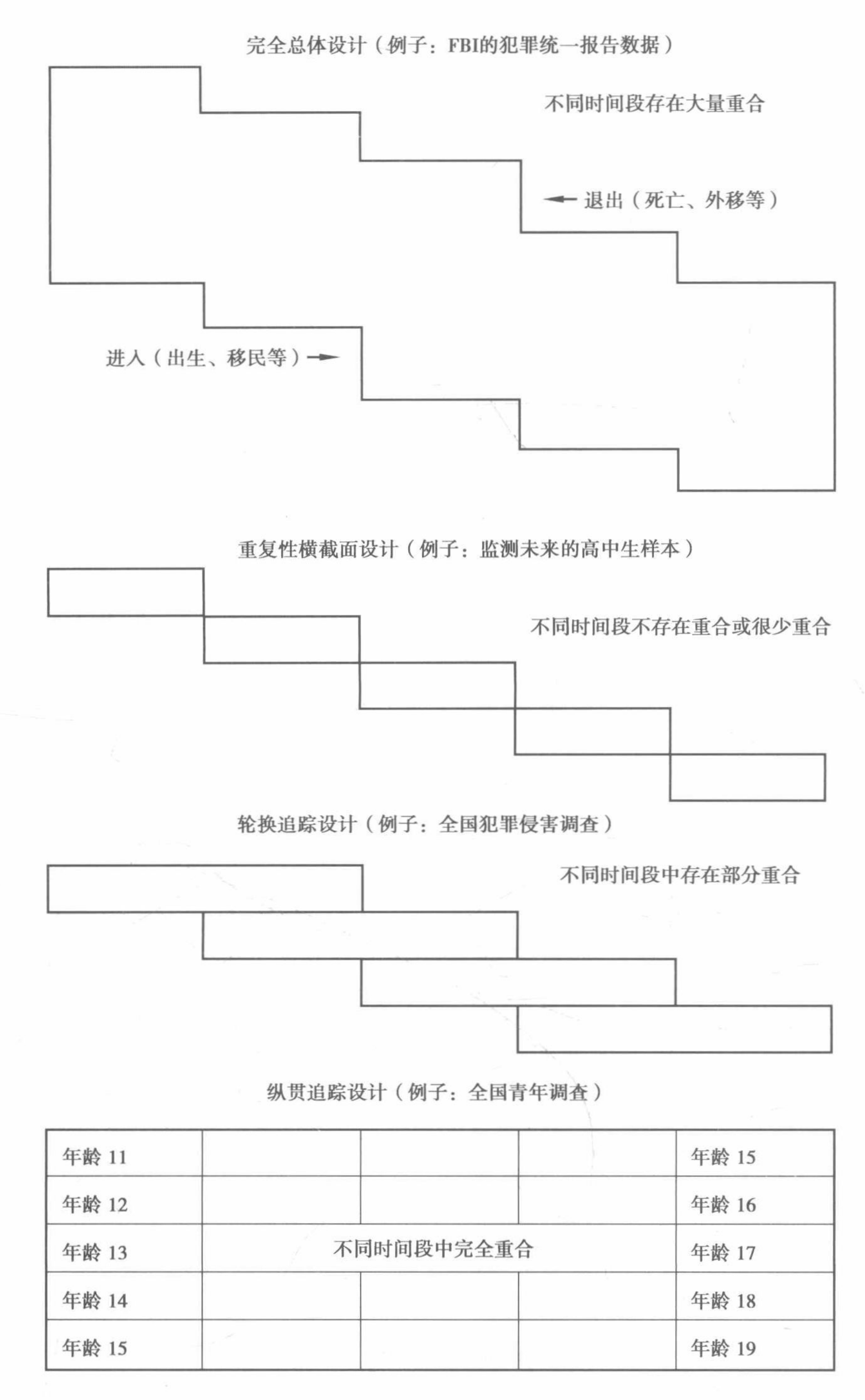


图 1　纵贯设计的不同类型

资料来源：Menard(2002)。

在完全总体设计（total population design）中，我们在不同时期收集整个总体的数据。完全总体设计的一个例子是联邦调查局数据的“犯罪统一报告（Uniform Crime Report）”，其中报告了警方所知的所有拘捕和犯罪记录。虽然覆盖率事实上达不到100%（十年一度的人口普查也不可能全覆盖），但是数据目标是包括全部美国人口。年复一年，人口进入总体中（出生），离开总体（死亡），所以数据在每一年与下一年中覆盖的个体略有不同但主体是相同的，但在长时间段中覆盖的个体就有差异了（如从1950年到现在）。FBI数据用来测量警方所知的拘捕和犯罪的变化率或趋势。收入动态追踪研究（Panel Study of Income Dynamics）也采用了相同的设计结构，这种设计虽然不是完全总体设计，但是其中的个体会因为死亡而离开样本，或者因为出生或婚姻而进入样本。

重复性横截面设计（repeated cross-sectional designs）在不同的时期收集不同样本的数据。因为每年抽取的都是新样本，原则上每一年与下一年的样本不存在重合部分（虽然独立样本也可能会抽取到同一个体）。重复横截面设计的最佳例子就是综合社会调查（General Social Survey），这是由全国民意研究中心（National Opinion Research Center）实施的一项年度综合调查，它涵盖了非常广泛的议题（包括婚姻和家庭、性行为与性角色、劳动力参与、教育、收入、宗教、政治、犯罪与暴力、健康、个人幸福），强调的是问题要完全相同，以便跨年度进行比较来测量宏观层面的变迁（Davis & Smith, 1992）。

轮换追踪设计（revolving panel designs）中会选取一组被访者在多个时期进行访谈，然后再更换为一组新的被访者。全国犯罪侵害调查（National Crime Victimization Survey, NCVS；司法年度统计局实施）就使用了轮换追踪设计。选取一些家户（household），在3年中采访7次：开始时一次，结束时一次，中间每6个月一次。在3年期结束时，会选择一户新家户来替换旧家户。替换是错开进行的，因此每6个月，NCVS调查中大约1/6的样本被替换。要注意的是，NCVS历史上就是家户样本，而不是个体样本：如果一个个体或一个家庭从这一户中搬走了，访谈的对象就不再是原来的被访者，而是原来家户中的（新）成员。因为三年后每个家户都会被替换，对于个体来说就只有短期的纵贯数据。但是，在全国层面上，NCVS是犯罪侵害率和犯罪率（限于找到受害者的犯罪）的宏观变迁上最广泛采用的数据来源之一。

纵贯追踪设计（longitudinal panel design）在不同学科中都被广泛承认是真正的纵贯设计。纵贯追踪设计的一个例子是全国青年调查（National Youth Survey, NYS）（参看Menard & Elliott, 1990）。NYS在1977年收集了前一年（1976年）中被访者的自报非法行为数据，包括吸毒行为。1977—1980年它每年都收集数据，1983—1992年它每3年收集数据，2003年时（即本文写作时）又对原初的被访者及家人收集了数据。不同于前面讨论的其他纵贯研究，第一年之后就没有人再进入样本。在最近几年中调查的被访者都是第一年调查到的被访者的子集：由于死亡、拒访、找不到等原因，有些原初的被访者没有接受调查。与综合社会调查、FBI数据、NCVS数据相比，NYS更少被用于考察犯罪的宏观历史变迁，而是更多用来考察个体的成长趋势，用来考察因果关系检验犯罪理论。在后一种情境中，NYS被用来研究各种侵犯事件及侵犯预测因素的发生时序，被用来构建只有用纵贯数据才能检验的跨时间因果模型【Causal】（Menard & Elliott, 1990）。

纵贯研究中可能遇到的问题

纵贯研究中可能会遇到横截面研究中的所有问题：内在和外在测量效度；测量信

度【Reliability】;抽样误差【Sampling Error】;拒访或不回答【Nonresponse】某些题项;问题对于研究总体来说是否合适;研究对象或被访者与访员、实验者、观察者之间的互动效应;研究成本。有些问题对于纵贯研究来说比横截面研究更加严重。例如,样本偏差在重复性横截面设计中会被放大,多年期纵贯调查要比单年横截面调查成本更大。此外,还存在一些其他危险。如前所述,如果需要回忆久远之前的事件,那么被访者的遗忘就会导致低估其非法行为。接受重复访问的被访者可能会了解到,某些回答会引发出一系列后续问题,因此会故意或无意识地避免回答这些问题带来的负担,这一问题也被称为追踪训练问题(panel conditioning)。与此相关,被访者或研究对象与实验者、访员、观察者之间的互动可能会导致无效回答,这一问题会因为前瞻性纵贯设计中实验者、访员或观察者与被访者或古代对象会重复进行接触而更加严重。在前瞻性纵贯设计的后期研究中,被访者可能已经死亡或者因为年龄或健康原因而不能接受访问,或者拒绝继续接受调查,或者研究者很难再找到这些被访者的居住地址,这都会导致追踪磨损(panel attrition)。在回溯性研究中,相应的问题也存在,那些本来应该包括在样本中以保证其代表性样本【Representative Sample】可能在研究开始前就已经故去或者已经找不到了。特别是在前瞻性纵贯性抽样调查【Survey】研究中,一个重要的问题是样本磨损是否非常有系统性或者是否非常大,以至于研究结论不能再推广到抽样基于的原初总体当中。

在纵贯研究开始时使用的测量可能在随后的研究中被认为已经过时,但是改变测量工具意味着数据不再有可比性。一个例子是1992年NCVS调查中提问方式发生了改变。新的方式导致了自报的受侵害率有了显著上升。用旧方法和新方法得出的受侵害率比较只能在单年中进行比较。因此,仍然难以确定那些"调整"受侵害率以使得旧方法与新方法之间具有对应关系的尝试是否成功,特别是在考察受侵害率的长期趋势时。在成长研究中,类似的问题是对同一个概念(如"亲社会联结")在生命历程的不同年龄采用不同的测量方式。例如,在青少年时期,这些联结主要发生在家庭取向的情境(如父母和兄弟姐妹)和学校情境中,但是在成年时期,这些联结主要发生在生育取向的家庭情境(如配偶和孩子)和工作情境中。研究者面临的是一个两难困境:要么提出对被访者年龄不相符的问题,要么使用不同的测量,在生命历程的不同阶段上的可比性就不能保证。

在横截面研究中,我们会有缺失数据【Missing Data】,因为个体会拒绝参与研究(研究对象缺失),或者会拒绝提供所需的全部信息(缺失值)。在纵贯研究中,还有一种可能性,就是个体会同意参与研究,在某一个或某一些时期也提供了所需数据,但是在另一个或另一些时期却拒绝提供或没有找到(缺失场合)(missing occasions)。分析纵贯数据的一些技术对于缺失数据的模式非常敏感,不能够处理长度不等的序列,因此即使数据中只是在一个变量上在一个场合中有缺失,也必须全部删除。纵贯研究中的缺失值问题可能有两种解决方式:一种是进行缺失值插补(用有根据的猜测来替换缺失值,猜测要基于其他数据或者研究中变量间的关系,有时包括了缺失情境的插补);另一种是使用那些分析中允许部分缺失数据的技术(Bijleveld & van der Kamp, 1998, pp.43-44)。

纵贯数据分析

纵贯研究让我们可以使用动态模型(dynamic models),即用一个或多个其他变量的变化来解决另一个变量的变化。虽然我们常常用横截面数据分析因果关系时的表述听起来似乎是在分析变化,但是我们其实只是在考察个体在一个变量上的差异如

何用这些个体在其他预测变量上的差异在解释。从个体间差异(interindividual differences)的发现中得到个体本身变化(intraindividual change)的结论,这已经是相对司空见惯的事情了。在特定情况下,可以用横截面数据来估计动态模型,但是这些条件是很苛刻的,很少能够在社会科学研究中满足(Schoenberg, 1977)。一般更合适的做法是使用纵贯数据,并且用能够发挥数据的纵贯性质的统计技术来分析数据。纵贯数据分析技术(Bijleveld & van der Kamp, 1998;Menard, 2002; Taris, 2000)包括:时间序列【Time-Series】分析可以在时期较多时用来描述和分析变迁(作为经济条件变化的函数的在押率变化);潜增长曲线模型【Growth Curve Model】和层级线性模型【Hierarchical Linear Model】可以用来描述和分析个体对象或案例的短期变迁和长期变迁,以及不同变量变化间的关系(如作为非法毒品使用的函数的暴力犯罪变化);事件史分析【Event History Analysis】可以用来分析对定类变量发生变化时机的影响(如学业成绩对于开始吸毒时间的影响)。

横截面研究不能够分解成长趋势和历史趋势,对历史变迁的描述和分析需要使用纵贯数据。使用横截面数据也可以尝试性地描述和分析成长趋势,但是结论与基于纵贯数据的结论可能不一定一致。只有用纵贯数据才能检验因果的时间顺序和序列问题。有一些横截面方法可以用来对互为因果关系(mutual causation)的模式进行建模,但是功能更强的模型需要用到纵贯数据,研究者可以用这些模型来更细致地考察因果关系,包括明显设定原因和结果的时间顺序。简言之,不存在用横截面数据可以作但用纵贯数据不可以作的分析,但是有许多分析只能用纵贯数据来作,而不能用横截面数据来作。横截面数据在描述某一特定时期的情况时仍然很有用,但是在社会科学中,纵贯数据越来越被认为是研究因果关系和历史变迁模式、成长变迁模式的最佳工具。

——Scott Menard

(高勇译校)

参考文献

Bijleveld, C. C. J. H., & van der Kamp, L. J. Th. (with Mooijaart, A., van der Kloot, W. A., van der Leeden, R., & van der Burg, E.). (1998). *Longitudinal data analysis: Designs, models, and methods*. London: Sage.

Davis, J. A., & Smith, T. W. (1992). *The NORC General Social Survey: A user's guide*. Newbury Park, CA: Sage.

Hill, M. S. (1992). *The panel study of income dynamics: A user's guide*. Newbury Park, CA: Sage.

Menard, S. (2002). *Longitudinal research*. Thousand Oaks, CA: Sage.

Menard, S., & Elliott, D. S. (1990). Longitudinal and cross-sectional data collection and analysis in the study of crime and delinquency. *Justice Quarterly*, *7*, 11-55.

Schoenberg, R. (1977). Dynamic models and cross-sectional data: The consequences of dynamic misspecification. *Social Science Research*, *6*, 133-144.

Taris, T. W. (2000). *A primer in longitudinal data analysis*. London: Sage.

局部加权回归(Lowess)

参见局部回归【Local Regression】。

M

宏观(Macro)

宏观是对集体层次的分析;微观是对个体层次的分析。因此,政治学家可以在宏观层面研究选举(如整个国家总统选举的结果),也可以在微观层面研究选举(如个体选民在总统票选中的抉择)。此外,经济学家可以研究宏观经济指标(如一国的国内生产总值),或者某特定产品(如汽车)供需的微观经济学。最后要指出的是,宏(macro)也是一种计算机程序子程序的名称,比如在SAS宏中。

——Michael S. Lewis-Beck

(李洋译　高勇校)

邮寄问卷(Mail Questionnaire)

如果可以得到调查总体(个体或者企业)的地址,那么邮寄问卷【Questionnaire】给概率样本【Probability Sample】来收集信息,就是一种成本上非常合算的方法。调查设计中既可以只使用邮寄问卷这一种方式,也可以采用自填问卷与其他填答模式(如电话、网络或面对面访谈【Interview】)相结合的复合方式。最常用的一种双模式设计是,先进行邮寄问卷,然后对没有返回邮寄问卷的样本进行电话访谈。多重模式设计可以减少无回答【Nonresponse】,从而增强得到的样本的代表性。

对于包括邮寄问卷的研究设计【Research Design】来说,抽样框【Sampling Frame】就是一组地址,比如公司的雇员列表,或者某一职业组织的成员列表。要注意的是,不能基于电话本中的地址列表对于总体进行抽样。因为有些家庭没有安装电话或者只使用手机,有些号码也没有列到电话本中,所以从电话列表中抽取的样本的代表性会因此而降低。

为了发现某些少见群体的成员,并对其进行邮件调查【Survey】,可以采用随机数字拨号【Random-Digit Dialing,RDD】技术。一旦发现一户中居住着某些少见群体的成员,访问员会询问其地址信息,然后向其邮寄问卷。使用RDD方法的优点是,每个有电话的家庭都有已知的被选中的机会,但是许多人不愿意在电话里向陌生人泄露自己的地址信息。

在邮寄问卷中,如何选择被访者是另一个挑战(比如,出现地址已知而姓名未知,或者需要父母随机选择一位子女进行回答的情形时)。有一些技术可以用来选定回答者或受试(如在代理回答【Proxy Reporting】的情况下),但是必须进行细致的设计,因此得

到的样本的质量取决于受访者是否遵从抽样要求。

研究者有许多方法可以用于开发调查工具,以便收集到能够充分解决其研究问题的数据。调查工具的开发通常从文献回顾开始,来发现解决手头问题已有的调查工具。接下来在目标总体中进行焦点小组【Focus Group】访谈,以发现重要的议题和人们用以讨论这些议题时所用的专用词汇。就这些内容草拟出备选的调查题项之后,还要与目标总体中的个体进行细致的、一对一的认知访谈。这些访谈的目标是发现那些人们在理解上存在不一致的调查问题和专用词汇。这样在调查实施之前,调查工具就得到了修正和检验。在调查实施前,量表通常还要经受专家的评估。相关领域的权威专家会提议内容,并评价调查工具是否适当。实施邮寄问卷前,最后一步是试调查【Pretest】。与目标总体特征一致的个体被召集在一起进行小组试调查。参与者在完成问卷之后,会被询问填答量表时的体验和回答问题时遇到的困难。小组试调查也可以估计出填答问卷表需要花费被调查者多长时间。

在决定是否适用邮寄问卷以及设计邮寄方式时,必须考虑目标总体的特征。为了减轻被调查者的负担,很重要的一点是要考虑问卷长度和阅读难度。在那些涉及弱势群体的研究(比如,那些阅读和写作能力较弱的人、视力不健全的人,或者健康状况较差、容易疲劳的人)中,问卷的长度和可读性尤其关键。

采用邮寄问卷时,现场没有访谈员可鼓励被调查者,或者回答他们提出的问题。因此,问题应该是易于回答的,跳答应该是易于进行的。封闭式问题【Closed-Ended Questions】只需要在框中画钩或涂黑圆圈(使用光学浏览技术进行数据录入)就行了,所以比开放式题项更可取;开放式题项不容易获取足够的回答,进行标准化编码【Coding】时也更麻烦。

使用邮寄问卷的方法具有一定的优势,包括对敏感话题能得到更好的估计。如果没有访谈员在场,人们更愿意报告那些被社会认为是不可取的行为(比如,药物滥用或暴力)。与电话或面对面访谈相比,另一个优势是能降低数据收集的成本。邮寄问卷还可以使调查工具避开看门人到达调查对象手中,被调查者会发现研究主题足够重要,值得花费时间进行回答。此外,由于被调查者不需要记住所有的回答选项,所以与电话调查相比,邮寄问卷中的回答选项可以更长。

邮寄研究的实施一般需要八个星期。尽管有许多公认的协议【Protocol】,但是标准邮寄调查研究经常包括三次邮件联系。首先,邮寄包括能说明研究目标和资助人的说明信、调查问卷、用来返还问卷已付邮资的信封的问卷包。一周之内,给所有样本成员邮寄用来提醒或感谢的明信片。首次邮寄之后的 2~3 周,给所有未回答的被调查者邮寄备用邮件包。回答率取决于联系信息的质量。在寄出的信封上印上核对地址的请求,这样能够得到那些已经搬家的样本成员的地址。

系统的质量控制措施是成功采集数据的关键。样本成员都会有一个标识号码附在问卷上面,这样便于查对。调查数据会采用手工或者光学扫描的方式进行录入,然后就可以进行统计分析了。

——Patricia M. Gallagher

(李洋译　高勇校)

参考文献

Dillman, D. A. (2000). *Mail and Internet surveys: The tailored design method.* New York: John Wiley.

Fowler, F. J., Jr. (2002). *Survey research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Schwarz, N. A., & Sudman, S. (Eds.). (1996). *Answering questions.* San Francisco: Jossey-Bass.

主效应(Main Effect)

主效应一词通常用于方差分析【Analysis of Variance】的析因设计【Factorial Design】中。在一个因子设计中,因子A的不同水平与因子B的不同水平有交互关系,如表1所示。

表1

		因子B			边缘均值
		水平 B_1	水平 B_2	水平 B_3	
因子A	水平 A_1	M_{A1B1}	M_{A1B2}	M_{A1B3}	M_{A1}
	水平 A_2	M_{A2B1}	M_{A2B2}	M_{A2B3}	M_{A2}
	边缘均值	M_{B1}	M_{B2}	M_{B2}	

每个单元格数据都是均值,边缘均值是一个水平的因子在其他因子的不同水平下取值的均值。例如,M_{A1}是在因子A的水平1下,因子B不同水平取值的均值。

通俗地讲,主效应【Main Effect】指的是因子的边缘均值的特征。研究者以不同方式使用这个词汇。有的研究者说"因子A的主效应存在",意思是指基于正式的统计检验结果,得出了总体【Population】中因子的边缘均值并非全都相等的结论。其他研究者则只是说"包含了因子A的主效应的均值",以此来强调这是边缘均值,但并不涉及它们的相等性或者对它们进行统计检验的结果。

有几种计算边缘均值的方法。一种方法使用未加权均值,边缘均值是在不同水平下的各个单元格均值的平均数。例如,M_{A1}定义为$(M_{A1B1}+M_{A1B2}+M_{A1B3})/3$。第二种方法建立在加权均值的基础上,每个纳入边缘均值计算的单元格均值,据其样本量【Sample Size】大小按比例加权。样本规模较大的均值在计算边缘均值时权重更大。

主效应也能用在线形回归【Linear Regression】模型中。"主效应"模型是指不包括任何交互【Interaction】项的模型。例如,"主效应"多元回归【Multiple Regression】模型是

$$Y = a + b_1X + b_2Z + e$$

而"交互模型(interaction model)"是

$$Y = a + b_1X + b_2Z + b_3XZ + e$$

第二个公式中的乘积项使其成为交互模型。在回归中,主效应可以用来称呼回归【Regression】模型中未被包括在乘积项或交互项中的任何一个连续的预测变量【Predictor】。主效应也可以用来称呼未被包括在乘积项或交互项中的任何一个虚拟变量【Dummy Variable】或一组虚拟变量。在下面的连续预测因子模型中,

$$Y = a + b_1X + b_2Z + b_3XZ + b_4Q + e$$

Q的系数可以被称为Q的"主效应",但是不能这样称呼X、Z和XZ的系数。

——James J. Jaccard

(李洋译　高勇校)

管理研究(Management Research)

在1960年代,管理学在美国迅速发展,从1980年代开始,在世界其他地方流行起来,并逐渐成为社会科学中相对年轻的新生力量。在早期,管理学较为倚重相关学科,如经济学、社会学、人类学、统计学和数学。但是越到最近,越来越多的人意识到,这些相关学科在处理管理领域中复杂的理论和实践问题时,其自身存在不足,难以应付。

管理研究中有三个鲜明的特征,这决定了管理研究既要采用传统社会科学方法,也要发现新的研究方法。第一,在管理实践和管理学术理论上存在显著的差异。管理实践是将各种理论和方法兼容并包来解决具体的问题;管理学术理论是要采用清晰的学科框架来审视管理的各个方面。管理研究吸纳了不同的学科,从高度定量的学科,如财政、经济和运筹研究,到高度定性的学科,如组织行为、人力资源发展,所以人们对何谓高质量研究很难达成共识。这也导致人们一直在争论单学科视角还是跨学科视角更具有相对优势(Keleman & Bansal, 2002; Tranfield & Starkey, 1998)。

第二,与管理研究中的利益相关人(stakeholders)有关,因为经理人是管理研究的假设研究对象,他能够控制着研究的调查通道【Access】,有权提供或撤销研究人员的经费资助。因此,管理研究几乎不能与政治和伦理问题割裂开来;同时研究者往往要根据经理的日程来调整具体的研究方法(Easterby-Smith, Thorpe, & Lowe, 2002)。因此,管理研究中几乎不可能采用扎根理论【Grounded Theory】支持者所提倡的随机应变的(emergent)抽样策略或理论抽样【Theoretical Sampling】策略,因为经理人可能会坚持提前协商好访谈计划;由于研究者占用雇员时间的机会成本问题,经理人也会严格限制研究的进入。所以,在大多数的情况下,都不得不对扎根理论的准则进行调适。

第三,管理实践需要想法和行动,二者缺一不可。大多数经理人不仅感到管理研究应该要导向实践结果,而且他们自己也能够在研究结果的指导下行事。因此,研究方法要么有付诸实践的潜力,要么要考虑研究者指导与否可能引发的实践结果。极端情况下,这导致纯粹的研究人员与行动研究者【Action Research】或顾问的疏离,前者试图超然于研究对象,后者从经验中学习并着手改变。这也引发出来许多研究方法,如参与式探究(Reason & Bradbury, 2000)和Mode Ⅱ研究(Huff, 2000)方法,这些方法可能将研究转化为管理实践。

当然,这三个特征也广泛存在于社会科学中,但是三个特征的并存驱动着管理研究在方法上进行改变和创新。在某些情况下,这导致了现有方法的调整和延伸;在另一些情况下,引发出了新的方法和程序。此外,由于管理和商业领域在学术界中的绝对规模,这些正在演进的方法有可能对社会科学产生较大的影响。在完成本文写作时,在英国大约有1/3的社会科学家在商学院工作。毫无疑问,类似的情况在美国有过之而无不及。

——Mark Easterby-Smith

(李洋译 高勇校)

参考文献

Easterby-Smith, M., Thorpe, R., & Lowe, A. (2002). *Management research: An introduction* (2nd ed.). London: Sage.

Huff, A. S. (2000). Changes in organizational knowledge production. *Academy of Management Review*, *25*(2), 288-293.

Keleman, M., & Bansal, P. (2002). The conventions of management research and their relevance to management practice. *British Journal of Management*, *13*, 97-108.

Locke, K. D. (2000). *Using grounded theory in management research*. London: Sage.

Reason, P., & Bradbury, H. (2000). *Handbook of action research: Participative inquiry and practice*. London: Sage.

Tranfield, D., & Starkey, K. (1998). The nature, organization and promotion of management research: Towards policy. *British Journal of Management*, *9*, 341-353.

多元协方差分析(MANCOVA)

参见多元协方差分析【Multivariate Analysis of Covariance】。

Mann-Whitney *U* 检验(Mann-Whitney *U* Test)

这一检验常用于确定两个独立的样本是否来自同一个总体。它被认为是最有效的非参数【Nonparametric】检验之一，适用于定序及以上层次数据类型。Mann-Whitney *U* 检验和 Wilcoxon 秩和检验(the Wilcoxon rank-sum test)实际上完全相同，因此各种参考文献中倾向于将它们看作一种检验，被称为 Wilcoxon-Mann-Whitney 检验。在不同参考书中实际的计算可能有差异，但结论是一致的。在样本无法满足正态性和方差齐性(homogeneity of variance)(Siegel & Castellan, 1988)条件下，它也可作 t 检验的一种替代选择。如果样本量小，可以通过专用表格，或者依靠基于计算机的统计软件包来计算概率大小。

这种检验建立在数据秩次(rank order)的基础上，目的是看两个潜在的总体分布是否相同。解决思路是，看一组样本的秩和与两组样本的秩的总体平均值的差异是否足够大，如果差异足够大，则可以表明两组样本并非来自同一个总体【Population】。从抽样分布【Population Distribution】中可知，从同一个总体抽取出来的任何一个样本，其秩和即使不完全相同，也都会近似于总体均值。"大"样本的秩和分布服从正态分布，总体均值【Mean】可以定义为

$$\mu = \frac{1}{2} n_1 (n_1 + n_2 + 1)$$

均值标准误为

$$SEM = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

有研究建议大样本应该是每组样本至少有 10 个；也有的认为最低为 25 个。在任何一种情况下，样本越大，下面的近似值越精确。显而易见，该检验建立在 z 检验【z-Test】的基础上

$$z = \frac{\bar{x} - \mu}{SEM}$$

将其中一个样本组的秩和 W_1 作为样本均值，采用 Wilcoxon 方法，上式变为

$$z = \frac{W_1 - \frac{1}{2}n_1(n_1 + n_2 + 1)}{\sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}}}$$

式中 W_1——两组样本中一组的秩和；

n_1——一组的样本量（≥10）；

n_2——另一组的样本量（≥10）。

因此，为了在样本数据中进行检测，第一步是对所有的原始数据进行排序，以序号变作为数据。如果出现分值“相等”的情况，分值相等的数据设定为秩相同。然后将 z 值与表中适当的数值进行比对，以确定 α 水平。

例如，表 1 的数据显示了两周之后，两组样本西班牙语测试的得分，每人采用同样的学习材料，但使用不同的学习方法。两组学生都是从一个二年级班级中随机抽样得来的。A 组采用教师引导讨论的方法，并以视频辅助教学，B 组采用计算机练习，学生自主学习。在这个案例中，得分分布呈现高度偏态，每个原始数据组的第二列显示了整个数据组（两组合并后）的秩。

问题是，如果将自然变差（natural variation）考虑进去，它们仍来自相同的总体吗？把每组的秩各自加起来，其中一个是 W_1。因此 z 的计算公式变成了

$$z = \frac{W_1 - \frac{1}{2}n_1(n_1 + n_2 + 1)}{\sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}}}$$

表 1　西班牙语测试的两组数据，且列秩次相加

	数据 A	秩 A	数据 B	秩 B
	86	2	92	1
	78	5	85	3
	75	8	80	4
	73	10.5	77	6
	73	10.5	76	7
	72	12	74	9
	70	14	71	13
	68	16	69	15
	64	17.5	64	17.5
	60	20	61	19
	55	22	57	21
	50	25	52	23
	46	26	51	24
$W=$		188.5		162.5
$n=$		13		13

注：相等秩取相邻秩的平均数。

$$z = \frac{188.5 - 1/2 \times 13 \times (13 + 13 + 1)}{\sqrt{\frac{13 \times 13 \times (13 + 13 + 1)}{12}}}$$

$$z = \frac{188.5 - 175.5}{\sqrt{380.25}} = \frac{13}{19.5} = 0.67$$

如果 α 为 0.05，存在显著性差异的 z 的最小取值是 1.96。因此，上述计算结果表明两组在秩分布上差异并不显著，看起来两个样本来自相同的总体。换句话说，至少在这个练习中，假设控制其他可能的额外影响变量（extraneous variables），学习方法不同并不会带来显著差异。

同样用 Mann-Whitney 法计算 U_1，即

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - W_1$$

然后，针对 U_1（所得值与上值相同）的 z 的计算公式是

$$z = \frac{U_1 - (n_1 n_2/2)}{\sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}}}$$

正如豪韦尔（Howell，2002）指出的，如果是简单线性关系，对任何样本组用这两种方法得到的结果，其差异都只是一个常数，因此如果需要的话，很容易从一种方法转换为另一种方法。

——Thomas R. Black

（李洋译　高勇校）

参考文献

Black, T. R. (1999). Doing Quantitative Research in the Social. Sciences: An Integrated Approach to Research Design, Measurement, and Statistics. London: Sage Publications.

Grimm, L. G. (1993). Statistical Applictions for the Behavioral Sciences. New York: John Wiley & Sons.

Howell, D. C. (2002). *Statistical Methods for Psychology*. (5th ed) Pacific Grove. CA: Duxbury.

Siegel, S., and Castellan, N. J. (1988). *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.

多元方差分析（MANOVA）

参见多元方差分析【Multivariate Analysis of Variance】。

边缘效应（Marginal Effects）

在回归【Regression】分析中，数据分析人员经常感兴趣的是解释并测量自变量【Independent Variable】（或解释变量）对因变量【Dependent Variable】（或回答）的效应。一种测量自变量效应的方法是计算它们的边缘效应。自变量的边缘效应测量了回归模型中自变量（如 X_i）的变化对因变量（如 Y）的期望变动的影响，尤其在自变量的变化极小时。

如果自变量是连续和可微的，那么自变

量 X 对因变量 Y 的边缘效应,可以通过引入关于 X 的偏导数 $E(Y|X)$ 计算出来。考察含两个自变量的线性回归模型

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

其中,Y_i 是第 i 个观测上的因变量取值,X_{1i} 是连续自变量的取值,X_{2i} 是二分变量的取值,或者等于 0 或者等于 1,β 是回归系数,ε_i 是误差项,假定误差项独立分布,均值为 0,且常方差(constant variance)为 σ^2。假设满足经典回归模型,因变量 Y 的条件均值是 $E(Y|X) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$。因此,在线性回归模型中,$X_1$ 对 Y 的期望变动的边缘效应,可以通过关于 X_1 的偏导数计算出来,也就是

$$\frac{\partial E(Y|X,\beta)}{\partial X_1} = \frac{\partial(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}{\partial X_1} = \beta_1 \qquad (1)$$

尽管如此,由于 X_2 是一个不可微的二分变量,因此 X_2 的边缘效应无法通过求偏导数计算出来。格林(Greene,2002)和朗(Long,1997)建议将 $E(Y|X,\beta)$ 的离散变化作为二分变量(此例子中的 X_2)从 0~1 的变化值,并设其他所有变量为恒量(通常是它们的均值),以此来对二分自变量的边缘效应进行解释。如下所示

$$\frac{\Delta E(Y|X,\beta)}{\Delta X_2} = (\beta_0 + \beta_1 X_1 + \beta_2 \times 1) - (\beta_0 + \beta_1 X_1 + \beta_2 \times 0) = \beta_2 \qquad (2)$$

从方程(1)和(2)中,我们得知线性回归模型自变量的边缘效应等于参数 β。因此,参数估计值(这里是 $\hat{\beta}_1$ 和 $\hat{\beta}_2$)可以这样进行描述:设其他所有变量为恒量,X_1 增加一个单位会导致因变量 Y 平均增加(或减少,这取决于 $\hat{\beta}$ 的符号)一个 $\hat{\beta}$。要注意的是,对二分变量来说,由于只能取值为 0 或 1,因此一个单位的变化等同于变量的取值范围。因而,可以这样描述虚拟变量【Dummy Variable】回归:设其他所有的变量为恒量,具有 X_1 的属性(相反则不具有该属性),意味着 Y 平均增长(或减少)一个 $\hat{\beta}$。

在非线性【Nonlinear】回归模型中,对自变量边缘效应的解释更为复杂,因为自变量一个单位变化的效应取决于所有自变量值和模型参数。结果,在非线性模型如 Logit、probit、多重 Logit、顺序 Probit、Logit 和事件计数模型(event count models)中,所公布的系数几乎不可能等于边缘效应。

考察一个二分选择模型(binary choice models),如 logit 模型或者 probit 模型

$$\Pr(Y = 1 | \mathbf{X}) = F(\mathbf{X\beta})$$

其中,连接函数(link function)F 是正态分布【Normal Distribution】或 Logistic 分布的累积分布。无论何种情况下,由于自变量的变化,自变量的边缘效应就是概率的期望变动。如果数据是可微的,边缘效应也可以通过求偏导数计算出来。例如,第 k 个自变量的边缘效应(X_k)是

$$\frac{\partial \Pr(Y = 1 | \mathbf{X},\boldsymbol{\beta})}{\partial X_k} = \frac{\partial \mathrm{F}(\mathbf{X\beta})}{\partial X_k} = f(\mathbf{XB})\beta_k \qquad (3)$$

其中,f 是标准正态分布的概率密度函数,或者是 Logistic 分布的概率密度函数,对所有观测量来说,$\mathbf{X}$ 是所有观测的自变量取值的矩阵,$\boldsymbol{\beta}$ 是参数向量。由于尺度因子(scale factor)$f(\mathbf{X\beta})$ 必定为非负的,方程(3)表明 X_k 的边缘效应的向度(或符号)完全取决于 β_k。尽管如此,其边缘效应的量级完全由 β 和 $f(\mathbf{X\beta})$ 的值共同决定。因而,要解释非线性模型中的边缘效应,就需要对因变量的性质(例如,连续的、二分的、定序的、定类的)、所分析的问题和研究人员选择的概率

分布等进行深入考察。关于边缘效应议题上的更多信息，可以参考 Long，1997；Greene，2002；Sobel，1995。

——Cheng-Lung Wang
（李洋译 高勇校）

参考文献

Greene, W. H. (2002). *Econometric analysis* (5th ed.). New York: Prentice Hall.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables.* Thousand Oaks, CA: Sage.

Sobel, M. E. (1995). Causal inference in the social and behavioral sciences. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 1-38). New York: Plenum.

边缘齐性（Marginal Homogeneity）

边缘齐性是多维列联表中的某些边缘分布相等的一种模型。在 $R \times R$ 表格中，如果对所有 i，都有单元格概率 π_{ij} 满足 $\sum_j \pi_{ij} = \sum_j \pi_{ji}$，即满足单变量边缘齐性。

在 $R \times R \times R$ 表格中，如果对于所有 (i,j) 来说，单元格概率 π_{ijk} 都满足 $\sum_k \pi_{ijk} = \sum_k \pi_{ikj} = \sum_k \pi_{kij}$，那么，它就满足双变量边缘齐性。

——Marcel A. Goom
（李洋译 高勇校）

边缘模型（Marginal Model）

对一组变量的一些（或全部）边缘分布以及最终的联合分布上增加约束条件（constraints）的统计模型称为边缘模型。在多元正态分布（multivariate inormal distribution）中，由于任意边缘分布的一阶矩和二阶矩都等于联合分布（joint distribution）中的一阶矩和二阶矩，因此边缘分布建模不会产生特殊的问题。对分类【Categorical】变量来说，联合分布和边缘分布的参数一般并不是一一对应的，因此需要用专门方法通过边缘模型来分析数据。

表 1 中显示了在三个时点上测量的响应变量 Y 的数据。

表 1 观测频次分布

	$Y_3=1$		$Y_3=2$	
	$Y_2=1$	$Y_2=2$	$Y_2=1$	$Y_2=2$
$Y_1=1$	121	11	41	44
$Y_1=2$	4	2	1	12

数据来源：密歇根大学政治和社会研究校际联盟（Inter-University Consortium for Political and Social Research，ICPS）的全国青年调查。

这里报告的数据是对 236 名男孩和女孩在使用大麻的问题上进行重复访谈的汇总数据。其中，我们使用分别在 1976 年 (Y_1)、1978 年 (Y_2) 和 1980 年 (Y_3) 收集的数据，回答选项编码为：1 = 没吸过大麻，2 = 吸过大麻。

$\pi_{123}(y_1, y_2, y_3)$ 表示 Y_1，Y_2 和 Y_3 相应的联合概率分布（joint probability distribution）；那么，Y_1 和 Y_2 的二元边缘分布是

$$\pi_{12}(y_1, y_2) = \sum_{y_3} \pi_{123}(y_1, y_2, y_3)$$

Y_1 的单变量边缘分布是

$$\pi_1(y_1) = \sum_{y_2} \sum_{y_3} \pi_{123}(y_1, y_2, y_3)$$

其他边缘分布也可以进行类似的定义。

假定 Y_2 给定时 Y_3 和 Y_1 是独立的，这一假定对三个变量的联合分布加了约束条件，这可以在初始三维列联表中通过对数线性模型【Log-Linear Model】$[Y_1Y_2, Y_1Y_3]$ 拟合，以此来检验上述模型。如果模型仅对联合分布加以限制条件，它就不是一个边缘模型。

尽管如此，与针对整个表格的结构提出假设相比，人们可能会对这样的假设感兴趣，即认为响应变量边缘分布的某些特征并不随着时间推移而改变。单变量边缘同质性假设认为 Y 的单变量边缘分布在三个时点上相同：

$\pi_1(y) = \pi_2(y) = \pi_3(y)$ 对于所有的 y 都成立。

在这个例子中，相应的观测单变量边缘效应分布见表 2。

表 2　单变量边缘分布

	Y_1	Y_2	Y_3
1	217	167	138
2	19	69	98
	236	236	236

在这些数据中，$L^2 = 82.042\,5$ 且 $df = 2\ (p<0.001)$，三种分布来自同一概率分布的假设并不成立。因此，在追踪研究中，有证据表明，随时间推移大麻使用呈现显著增长。要注意的是，由于三种分布并不是建立在独立样本的基础上，这个假设难以通过传统的独立性卡方检验【Chi-Square Test】来进行检验。

或者是人们可能会对某些二分边缘分布相同的假设感兴趣，比如，对于所有的数据对 (y_1, y_2) 来说，都有

$$\pi_{12}(y_1, y_2) = \pi_{23}(y_1, y_2)$$

(Y_1, Y_2) 和 (Y_2, Y_3) 相应的观测二元分布见表 3 和表 4。

表 3　二元边缘分布 (Y_1, Y_2)

	$Y_2 = 1$	$Y_2 = 2$
$Y_1 = 1$	162	55
$Y_1 = 2$	5	14

表 4　二元边缘分布 (Y_2, Y_3)

	$Y_3 = 1$	$Y_3 = 2$
$Y_2 = 1$	125	42
$Y_2 = 2$	13	56

由于 $L^2 = 86.002\,4$ 且 $df = 4\ (p<0.001)$，因此两个边缘分布来自同一个二分概率分布的假设也被拒绝。

限制较少的假设是每个二元分布的相应关联参数相等。假设二元分布 π_{12} 和 π_{23} 中相应的局部优势比【Odds Ratios】相等，可以得到下面类型的限制条件，即对于所有的取值对 (y_1, y_2)，都有

$$\frac{\pi_{12}(y_1, y_2)\pi_{12}(y_1+1, y_2+1)}{\pi_{12}(y_1+1, y_2)\pi_{12}(y_1, y_2+1)} = \frac{\pi_{23}(y_1, y_2)\pi_{23}(y_1+1, y_2+1)}{\pi_{23}(y_1+1, y_2)\pi_{23}(y_1, y_2+1)}$$

在一个 2 × 2 的表格中，只需要考虑一个优势比。对于当前的数据来说，在(Y_1，Y_2)边缘表中，优势比等于 8.247；在(Y_2，Y_3)边缘表中，优势比等于 12.820 5。由于 $L^2 = 0.456\ 1$，且 $df = 1$ ($p = 0.499$)，因此无法拒绝两个优势比相等的假设。两个边缘表格的恒量优势比的估计值是 11.346 3。这表明在时间上接近的两个响应之间存在着较强的正向、持续关联。对于有两个以上的响应类别的变量来说，可以建构起与对应的相邻优势比(adjacent odds ratios)、累积优势比(cumulative odds ratios)或全域优势比(global odds ratios)相等的备择边缘模型(alternative marginal models)。

最后，也可以对根据边缘界定的条件分布加以限制条件。例如，假设给定 Y_1 时 Y_2 的条件分布，等于给定 Y_2 时 Y_3 的条件分布，那么就会有下列限制条件：对于所有的数据对(y_1，y_2)来说都有

$$\frac{\pi_{12}(y_1, y_2)}{\pi_1(y_1)} = \frac{\pi_{23}(y_1, y_2)}{\pi_2(y_1)}$$

结合当前的数据，可以发现

$$\Pr(Y_2 = 2 \mid Y_1 = 1) = 0.253\ 5$$
$$\Pr(Y_2 = 2 \mid Y_1 = 2) = 0.736\ 8$$
$$\Pr(Y_3 = 2 \mid Y_2 = 1) = 0.251\ 5$$
$$\Pr(Y_3 = 2 \mid Y_2 = 2) = 0.811\ 6$$

由于 $L^2 = 0.501\ 6$，$df = 2$ ($p = 0.778$)，无法拒绝相应转移率【Transition Rate】随时间推移趋向恒量这一假设。这些恒量转移概率的估计值是：

$$\Pr(Y_{t+1} = 2 \mid Y_t = 1) = 0.252\ 9$$

和

$$\Pr(Y_{t+1} = 2 \mid Y_t = 2) = 0.795\ 9$$

这些结果表明在一个特殊时间点上的大约 80%的使用者将会在下一个时间点继续原有做法；反之，在同一个时期，大约 25%的非使用者会成为使用者。

对边缘模型进行参数估计，并检验模型是否能实现数据拟合，这都需要专门设计的方法。在许多案例中，边缘模型可以建构成广义对数线性模型，后者已经发展出了适用的估计和检验方法。这些方法实际上是在对联合分布和边缘分布加以限制条件的情况下，对联合分布概率进行的最大似然估计【Maximum Likelihood Estimation】，同时可以对假设模型进行条件和非条件检验(unconditional tests)。

——Marcel A. Croon

(李洋译　高勇校)

参考文献

Bergsma, W. (1997). *Marginal models for categorical data*. Tilburg, The Netherlands: Tilburg University Press.

Croon, M. A., Bergsma, W., & Hagenaars, J. A. (2000). Analyzing change in categorical variables by generalized log-linear models. *Sociological Methods & Research*, *29*, 195-229.

Vermunt, J. K., Rodrigo, M. F., & Ato-Garcia, M. (2001). Modeling joint and marginal distributions in the analysis of categorical panel data. *Sociological Methods & Research*, *30*, 170-196.

边缘(Marginals)

如果多元联合分布(multivariate joint distribution)沿一个变量或多个变量维度进

行压缩(collapsing),就可以得到这个分布的边缘或边缘分布(marginal distributions)。如果 X,Y 和 Z 是概率分布为 $\pi_{XYZ}(x,y,z)$ 的三个离散随机变量(discrete random variables),X 和 Y 的边缘分布是

$$\pi_{XY}(x,y)=\sum_{z}\pi_{XYZ}(x,y,z)$$

X 的边缘分布是

$$\pi_{X}(x)=\sum_{y}\sum_{z}\pi_{XYZ}(x,y,z)$$

对于连续变量,加总变成积分即可。

——Marcel A. Croon

(李洋译 高勇校)

马尔科夫链(Markov Chain)

马尔科夫链描述了事物改变状态的概率。它关注的是在一个离散时间周期内改变状态的概率,例如从偏好政党 A 到偏好其他政党。这是一个不含记忆的过程:在当前测量时点下处于当前状态的概率,仅取决于上一时点的状态。更早时点的状态不影响预测值。一旦我们观察到两个相继时点的交互表,我们就可以对状态空间中所有组合的转换概率进行估计:当前状态为 A,下一时点中继续为 A 的概率为 $\tau_{A|A}$;当前状态为 A,下一时点转变为状态 O 的概率为 $\tau_{O|A}$;$\tau_{A|O}$ 和 $\tau_{O|O}$ 也与此类似。多状态时的情况很容易由此推出。

马尔科夫链是俄国数学家安德雷·马尔科夫提出来的。令人感兴趣的是,马尔科夫链可以对未来的状态分布进行预测。以表 1 的数据为例,很容易证明政党 A 刚开始有 67%的支持率,但是一年之后支持率降低了 5%。最终马尔科夫链将会实现均衡,政党 A 得到总体 58%的支持,这与状态的原始分布情况无关。这可以通过矩阵乘法计算得出,即矩阵本身乘以转换概率矩阵 T,其结果再乘以 T,直到结果矩阵的各行都相等。

如果有包含三个及以上时刻的定群【Panel】数据,那么就可以对马尔科夫链的拟合度进行检验。这样,马尔科夫链就可以在大多数社会科学应用中对很长时期的变化进行预测。布吕芒、科根和麦卡锡(Blumen, Kogan, & McCarthy, 1968)假定了两种类型的人群混合在一起:一些人改变状态,另一些人不改变状态。在三个及以上的时点上进行数据的定群观测,便可以估计出“迁移者”和“留守者”的比例。

表 1 政治倾向状态的改变

初始状态	一年以后;频次		一年以后;转换概率		马尔科夫链条件下的 7 年转换概率	
	政党 A	其他	政党 A	其他	政党 A	其他
政党 A	750	250	0.75	0.25	0.58	0.42
其他	175	325	0.35	0.65	0.58	0.42

对变迁进行分析时,陷阱之一在于测量误差会使得最终结果不精确。即使时点之间是相互独立的,这些误差仍会放大观测到的变化,除非将这些误差项纳入模型中。潜马尔科夫模型【Latent Markov Model】(也被称为隐马尔科夫模型)允许结果中包括误差。状态空间中的每个回答都可能出现,但是其概率取决于当前状态。人们能进行正确回答的概率很高,而回答错误的概率虽然较低但仍然存在。通过三个及以上时点的定群观测,就可以对这个模型进行估计。

这里可以将两种方法结合起来,例如,假设总体中的一部分呈现出潜马尔科夫链的特征,另一部分由稳定者(stayers)组成。为了对这个局部混合的潜马尔科夫模型进行准确的估计,最好在每个时点都引入多重测量指标。

如果人们倾向于返回之前的状态,这种情况就比较复杂,如劳动力市场状况研究中的失业问题。这时,可以将先前状态作为一个影响失业概率的外生变量来建模。

最后,离散时间转换概率实际上是更基本的连续时间流动参数的结果——转换率【Transition Rates】或者风险率【Hazard Rates】——在无穷小的时间周期中的应用。

——Frank van de Pol

Rolf Langeheine

(李洋译　高勇校)

参考文献

Blumen, I., Kogan, M., & McCarthy, P. J. (1968). Probability models for mobility. In P. F. Lazarsfeld & N.W. Henry (Eds.), *Readings in mathematical social science* (pp. 318-334). Cambridge: MIT Press.

Langeheine, R., & van de Pol, F. (1990). A unifying framework for Markov modeling in discrete space and discrete time. *Sociological Methods & Research*, *18*, 416-441.

Vermunt, J. K., Rodrigo, M. F., & Ato-Garcia, M. (2001). Modeling joint and marginal distributions in the analysis of categorical panel data. *Sociological Methods & Research*, *30*, 170-196.

马尔科夫链蒙特卡洛法(Markov Chain Monte Carlo Methods)

马尔科夫链蒙特卡洛(MCMC)法是一种模拟【Simulation】技术,可以通过建构平稳分布(stationary distribution)为 p 的马尔科夫链【Markov Chain】,并对马尔科夫链进行多次迭代,以得到近似服从概率分布 p 的随机样本。MCMC 法应用最多的领域是贝叶斯推论【Bayesian Inference】。这里,通过 MCMC 法,可以通过数值计算对贝叶斯后验分布【Posterior Distribution】进行归纳,这个结果无法用解析解的方法求得。

在许多应用中,研究人员发现能从任意概率分布 p 中进行抽样很有用。当人们希望计算某个积分的数值近似值时这一点尤其有用,比如在贝叶斯推理以及似然推理中就会是这样。当 p 属于一个已知分布族,有时候采用标准伪随机数生成技术(pseudo-random number generation techniques)来直接从 p 中生成一串独立的伪随机离差是可能的。但是在许多应用实践中,p 并不属于已知的分布族,无法使用标准方法直接从中进行抽样。在这种情形下可以使用 MCMC 法,而且操作起来也相对容易。

MCMC 法并非直接从 p 中生成一个独立样本,而是通过从特定的随机过程中相继

抽取一个连续数列来发挥作用,这个随机过程的长期均衡分布(long-run equilibrium distribution)为 p。通过建构一个足够长的数列,我们可以将靠后的抽取看作近似于服从分布 p。更准确地说,MCMC 法提供了建构马尔科夫链的方法,后者的均衡(或平稳)分布就是感兴趣的分布 p。

由于 MCMC 法所抽取的数列不是直接来自 p 的独立数列,因此在使用 MCMC 法时会有许多复杂问题。其中之一涉及要抛弃多少初始的所谓测试迭代(burn-in iterations)。原因在于,MCMC 抽样在马尔科夫链运行很长一段时间之后才能近似于服从 p 分布,所以初始的 MCMC 抽取结果并不能代表从分布 p 中抽取的结果,故而应当舍弃。另一个复杂的问题体是要确定马尔科夫链需要运行的长度,以便得到一个样本,使之充分接近于在使用中直接从 p 中进行的抽样。这些问题或多或少的在标准参考文本中,如罗伯特和卡塞拉(Robert and Casella,1999)有所论述。

研究人员通常使用两种基本的 MCMC 算法(尽管要注意其中一种类型实际上是另一种的特例)。第一种被称为 Metropolis-Hasting 算法;第二种也属于 Metropolis-Hasting 算法的一类,被称为 Gibbs 抽样【Gibbs Sampling】算法。

Metropolis-Hasting 算法

Metropolis-Hasting 算法是一种很普通的 MCMC 算法,可以用它从仅知道其加总等于标准化常数的概率分布中进行抽样。Metropolis-Hasting 算法易于从进行模拟的候选生成密度(candidate-generating density)中生成候选值,进而概率性地接受或者拒绝这些候选值,以产生一个近似于来自分布 p 的样本。

假设我们要从 $p(\theta)$ 中随机抽取一个样本,其中 θ 是多维的。图 1 对 Metropolis-Hasting 算法进行了展示。

initialize $\theta^{(0)}$

for (i in 1 to m) {

generate $\theta_{can}^{(i)}$ from $q(\theta|\theta^{(i-1)})$

calculate $\alpha \leftarrow \dfrac{f(\theta_{can}^{(i)})}{f(\theta^{(i-1)})}\dfrac{q(\theta^{(i-1)}|\theta_{can}^{(i)})}{q(\theta_{can}^{(i-1)}|\theta^{(i-1)})}$

set

$$\theta^{(i)} \leftarrow \begin{cases} \theta_{can}^{(i)} \text{ with probability } \min\{1,\alpha\} \\ \theta^{(i-1)} \text{ with probability } 1-\min\{1,\alpha\} \end{cases}$$

store $\theta^{(i)}$

}

图 1

其中,$f(\theta)$ 是与 $p(\theta)$ 成比例关系的一个函数,$q(\theta|\theta^{(i-1)})$ 是备选生成密度。要注意的是,备选生成密度取决于 $\theta^{(i-1)}$ 的当前值。当备选生成密度对称时——也就是 $q(\theta|\theta^{(i-1)})$ 等于 $q(\theta^{(i-1)}|\theta)$——那么这些项在接受概率 α 的公式中就消掉了。

Gibbs 抽样算法

Gibbs 抽样算法是 Metropolis-Hastings 算法在接受概率(acceptance probability)总等于 1 时的一种特例。它通过从 p 的条件分布中进行迭代抽样,生成一个近似于来自联合分布 p 的样本。

为了考察 Gibbs 抽样算法是如何运作的,假设我们有一个联合分布 $p(\theta_1,\cdots,\theta_k)$,将对 p 的变元(arguments)分解成 k 个不同的区组(block)。这里用 $p_1(\theta_1|\theta_2,\cdots,\theta_k)$ 表示 θ_1 的条件分布,且剩余的 $k-1$ 个区组与其他 $k-1$ 个完全条件分布的标记相同。图 2 显示了 Gibbs 抽样算法的运算过程。

```
initialize θ_2^(0), …, θ_k^(0)
for (i in 1 to m) {
    generate θ_1^(i) from p_1(θ_1 | θ_2^(i-1), …, θ_k^(i-1))
    generate θ_2^(i) from p_2(θ_2 | θ_1^(i), θ_3^(i-1), …, θ_k^(i-1))
    ⋮
    generate θ_k^(i) from p_k(θ_k | θ_1^(i), …, θ_{k-1}^(i))
    store θ_1^(i), …, θ_k^(i)
}
```

图 2

历史沿革

MCMC 法起源于统计物理。梅特罗波利斯等人（Metropolis, Rosenbluth, Rosenbluth, Teller and Teller, 1953）提出了现在众所周知的 Metropolis 算法来进行数值积分（numerical integrations）运算，以探索物理系统的特定性质。黑斯廷斯（Hastings, 1970）概括了 Metropolis 算法，引入非对称的备选生成密度进行运算。S. 杰曼和 J. 杰曼（Geman & Geman, 1984）在图像复原问题上，提出了现在称为 Gibbs 抽样算法的算法。尽管有上述这些（以及其他几个）较早期的工作，但是统计学家们并没有完全理解 MCMC 法的潜力，直到盖尔范德和史密斯（Gelfand and Smith, 1990）提出如何使用 MCMC 技术，来对在当时看来计算上不可行的模型进行贝叶斯推理，人们才对其应用潜力有了更好的理解。

举例：贝叶斯后验分布抽样

考察简单 Logistic 回归【Logistic Regression】模型的贝叶斯处理。假设我们有 n 个观测，第 i 个观测的因变量的值（y_i）编码为 1 或者 0。假定人们之前认为系数向量 β 服从正态分布，均值为 b_0，方差协方差矩阵为 B_0，则 β 的后验密度是（最多一个比例常数）

$$p(\beta \mid y) \propto \prod_{i=1}^{n}\left(\frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)}\right)^{y_i} \times \left(1-\frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)}\right)^{1-y_i} \times \exp\left(-\frac{(\beta-b_0)'B_0^{-1}(\beta-b_0)}{2}\right)$$

为了对后验密度进行概括，我们可能需要计算出 β 的后验均值和后验方差等。假设我们要用 β 的后验均值［定义为 $E[\beta \mid y]=\int \beta p(\beta \mid y)\mathrm{d}\beta$］进行贝叶斯点估计。即使对于一个简单的模型来说，进行必要的积分运算来计算后验均值也不是一个简单的问题。另一个对 β 的后验均值进行计算的方法，是采用 MCMC 算法，从 $p(\beta|y)$ 中抽取大小近似为 m 的随机大样本 $\beta^{(1)},\cdots,\beta^{(m)}$，并将 β 的后验均值估计为

$$\hat{E}[\beta \mid y]=\frac{1}{m}\sum_{i=1}^{m}\beta^{(i)}$$

下面的图 3 显示了 Metropolis 算法的运算过程。

```
initialize β^(0)
for (i in 1 to m) {
    generate β_can^(i) ← β^(i-1) + z^i
    calculate α ← f(β_can^(i) | y) / f(β^(i-1) | y)
    set β^(i) ←
    { β_can^(i) with probability min{1, α}
    { β^(i-1) with probability 1−min{1, α}
    store β^(i)
}
calculate β̄ ← (1/m) Σ_{i=1}^{m} β^(i)
```

图 3

其中 $z^{(i)}$ 是一个均值为零的多元正态随机变量，其固定方差协方差矩阵 $f(\beta|y)$ 是在第 i 次迭代中得到的，而 $f(\beta|y)$ 是与 $p(\beta|y)$ 成

比例关系的任意函数。要注意的是,实际上,一般要将 β 最初几百到几千个样本作为测试迭代(burn-in iterations)予以抛弃,以消除对 β 初始值的敏感性。

——Kevin M. Quinn

(李洋译 高勇校)

参考文献

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*, 398-409.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 721-741.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains. *Biometrika*, *57*, 97-109.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculation by fast computing machines. *Journal of Chemical Physics*, *21*, 1087-1092.

Robert, C. P., & Casella, G. (1999). *Monte Carlo statistical methods.* New York: Springer.

匹配(Matching)

匹配,是指基于在一个或多个协变量上的相似性,进行试验者【Treatment】和对照者【Control】的配对。在随机实验中,匹配减少了误差方差,因此能增强统计检力【Statistical Power】。在非随机实验中,匹配能增强实验组和控制组的相似性。非随机实验中的匹配并不能像随机化那样使各组完全均等,但它能减少偏差。

匹配包括以下类型。成对匹配(matched-pairs)是指一个试验者匹配一个对照者,由此得到的试验组和对照组大小相同。成组匹配(Matched-sets)是指一个或多个试验者匹配一个或多个对照者。匹配可以基于单个变量(单变量匹配)或多个变量(多元【Multivariate】匹配)。精确匹配(exact matching)是指将协变量【Covariate】取值完全相同的试验者和对照者配对;但是给所有对象都找到精确匹配可能太困难了。精确多元匹配更易于办到,方法是将匹配变量压缩(collapsing)成一个称为倾向值(propensity score)的变量,它是从 Logistic 回归【Logistic Regression】中得到的属于实验组还是对照组的预期概率。距离匹配(distance matching)进行配对的原则是使对照组和实验组之间的总体差异最小化。准确和距离匹配的其他形式有全匹配、最优匹配(optimal matching)和贪婪匹配(greedy matching)(Shadish, Cook, & Campbell, 2002)。

匹配的相关研究案例主要包括:(a)伯格纳(Bergner,1974)在关于夫妻沟通测量中,根据前测得分对 20 对夫妻进行排序,然后从每对中随机安排一对夫妻进入实验组或对照组;(b)奇切雷利和阿苏什埃兹(Cicerelli & Associates,1969)对加入"启蒙教育计划(Head Start)"的儿童,与同样性别、种族和幼儿园教育的、有义务加入但是没有加入的儿童进行比较;(c)德赫贾和沃赫拜(Dehejia & Wahba,1999)对接受劳动培训的试验组和从全国调查中抽取的控制组进行倾向值匹配。如果不谨慎从事,在非随机实验中进行匹配可能引发严重的问题(Shadish et al., 2002)。

——William R. Shadish

M. H. Clark

(李洋译 高勇校)

参考文献

Bergner, R. M. (1974). The development and evaluation of a training videotape for the resolution of marital conflict. *Dissertation Abstracts International*, *34*, 3485B. (UMI No.73-32510)

Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts.* New York: Guilford.

Cicerelli, V. G., & Associates. (1969). *The impact of Head Start: An evaluation of the effects of Head Start on children's cognitive and affective development* (2 vols.). Athens: Ohio University and Westinghouse Learning Corp.

Dehejia, R. H., & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, *94*, 1053-1062.

Rosenbaum, P. R. (1998). Multivariate matching methods. *Encyclopedia of Statistical Sciences*, *2*, 435-438.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for general causal inference.* Boston: Houghton Mifflin.

矩阵(Matrix)

矩阵(matrix)可以定义为由排成行和列的真实的数字或元素所组成的矩形排列。矩阵的维度或阶数由行和列的个数决定。比如,说 **A** 是一个 m 乘 n(表示为 $m \times n$)的矩阵,是因为它包含了 $m \times n$ 个排成 m 行和 n 列的真实的数字或元素,比如下面的矩阵

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & a_{ij} & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

其中,$a_{ij}(i=1,2,\cdots,m;j=1,2,\cdots,n)$ 是矩阵 **A** 的一个元素【Element】。

要注意从左上角开始算起,矩阵元素 a_{ij}的第一个下标表示行数,第二个下标表示列数。比如,a_{ij}是位于第 i 行和第 j 列的元素取值。如果矩阵 **B** 中对于所有 i 和 j 来说都有 $a_{ij}=a_{ji}$,那么 **B** 就被称作对称矩阵(symmetric matrix)。比如,

$$\mathbf{B} = \begin{bmatrix} 1 & 4 & 7 \\ 4 & 6 & 5 \\ 7 & 5 & 9 \end{bmatrix}$$

由于只有一列,$m \times 1$ 矩阵被称为列向量(column vector)。同样,由于只有一行,$1 \times n$矩阵被称为行向量(row vector)。1×1 矩阵仅是一个数,通常称作标量(scalar)。

行数和列数相同的矩阵叫作方阵(square matrix)。比如,上面图示的对称矩阵 **B** 是一个 3×3 的方阵。方阵对角线上元素之和是这个方阵的迹(trace)。例如,迹(**B**)= 16。几种特殊类型的矩阵常常被用在统计分析和计量经济学【Econometric】分析中。

- 对角线矩阵(diagonal matrix)是位于主要(主)对角线(即从左上角到右下角)之外的每个元素都等于零的方阵。即,如果 $i \neq j, a_{ij}=0$。例如,**C** 就是一个 $m \times m$ 的对角线矩阵,

$$\mathbf{C} = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & & 0 \\ 0 & 0 & 0 & a_{mm} \end{bmatrix}$$

• 等值矩阵(identity matrix)或单元矩阵是一种特殊的对角线矩阵,其主对角线上的元素全都等于1。通常表示为$\mathbf{I}_m$,其中m是单位矩阵的维度或阶乘[因而,迹($\mathbf{I}_m = m$)]。例如,$\mathbf{I}_3$就是一个等值矩阵,

$$\mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

• 三角矩阵(triangular matrix)是在主对角线上方或下方只有数字零的特殊方阵。如果零在对角线以上,矩阵就是下三角(lower triangular)矩阵,因为非零元素都在对角线以下。也就是

$$a_{ij} = \begin{cases} a_{ij}, i \geqslant j \\ 0, i < j \end{cases}$$

例如,矩阵$\mathbf{D}$是一个下三角矩阵,

$$\mathbf{D} = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & & 0 \\ a_{m1} & a_{m2} & 0 & a_{mm} \end{bmatrix}$$

相反,当零都位于对角线以下,矩阵就是上三角(upper triangular)矩阵。

• 主对角线上的元素全部相同的对角线矩阵称作标量矩阵(scalar matrix)。比如,满足经典线形回归【Regression】分析所有假设的线性回归【Linear Regression】误差变量,其方差-协方差矩阵如下所示:

$$\Omega = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & & 0 \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

这些矩阵中的任一矩阵均有其自身的特征,通常用在统计学【Statistics】和计量经济学中。比如,矩阵运算以及矩阵的转置(transpose)和求逆是统计分析和数理经济运算中的有效工具(关于这些议题的更多资料,基础论述可以参考 Namboodiri 1984, Chiang, 1984;高级论述可以参考 Golub & Van Loan, 1996)。

——Cheng-Lung Wang

(李洋译　高勇校)

参考文献

Chiang, A. C. (1984). *Fundamental methods of mathematical economics* (3rd ed.). New York: McGraw-Hill.

Golub, G. H., & Van Loan, G. F. (1996). *Matrix computations* (3rd ed.). Baltimore: Johns Hopkins University Press.

Namboodiri, K. (1984). *Matrix algebra: An introduction.* Beverly Hills, CA: Sage.

矩阵代数(Matrix Algebra)

矩阵操作能用于对多个变量在多次观测中的复杂关系进行求解。除非极其简单,矩阵操作一般都通过计算机程序来执行,但是三种基本的矩阵运算【Matrix】是共通的,并且易于说明。

矩阵加法

阶乘为$m \times n$的矩阵$\mathbf{A} = [a_{ij}]$和$\mathbf{B} = [b_{ij}]$,相加后得到矩阵$\mathbf{A}+\mathbf{B}$,阶乘为$m \times n$,也就是$\mathbf{A}+\mathbf{B} = [a_{ij}+b_{ij}]$。例如,如果

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

且

$$\mathbf{B} = \begin{bmatrix} e & f \\ g & h \end{bmatrix}$$

那么

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} a+e & b+f \\ c+g & d+h \end{bmatrix}$$

交换率和结合率在矩阵加法中同样都有效。

乘以一个数

如果 $\mathbf{A} = [a_{ij}]$ 是一个矩阵,且 k 是一个数字,那么矩阵 $k\mathbf{A} = [ka_{ij}] = \mathbf{A}k$。例如,如果

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

那么

$$k\mathbf{A} = \begin{bmatrix} ka & kb \\ kc & kd \end{bmatrix}$$

要注意的是矩阵的每个元素都要乘以一个数(与行列式乘以一个数不同)。乘以一个数字是满足交换率的。

矩阵乘法

只有两个矩阵可相乘(conformable)时,两个矩阵才可以相乘。可相乘意味着第一个矩阵的列数必须等于第二个矩阵的行数。矩阵 $\mathbf{A} = [a_{ik}]$,阶乘为 $m \times p$,矩阵 $\mathbf{B} = [b_{kj}]$,阶乘为 $p \times n$,阶乘为 $m \times n$ 的矩阵 $\mathbf{A} \times \mathbf{B}$是

$$\mathbf{AB} = \left[\sum_{k=1}^{p} a_{ik} b_{kj}\right]$$

由于矩阵必须可相乘,因此相乘的顺序也会产生差异(即,矩阵乘法不满足交换率)。对相乘得到的矩阵 **AB**,我们说 **A** 左乘 **B**,或者 **B** 右乘 **A**。对于此例,当 $m = p = n = 2$ 时,可以得到

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

以及

$$\mathbf{AB} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}$$

可以看出,**AB** 的每个元素都是 **A** 的特定元素和 **B** 的特定元素的乘积和。还可以看出,每个乘积对子的内标(the inside number of each product pair)都是相等(因为都为 k),而且在加总时这一内标数字从 1 升到 p(在这个例子中等于 2)。如果两个矩阵规模不同,则乘积矩阵 **AB** 和 **BA** 的阶数可能与初始矩阵的阶数并不一致,而且 **AB** 和 **BA** 的阶数也可能彼此不同。还有可能其中一个根本不存在。例如,如果有 $\mathbf{A}_{2\times 4}$ 和 $\mathbf{B}_{4\times 1}$,那么就有 $\mathbf{AB}_{2\times 1}$,但是 **BA** 就不存在,因为 **B** 的列数与 **A** 的行数不相等。要注意,即使 **AB** 和 **BA** 同时存在且矩阵规模相同,这两个矩阵也不必然相等。

有了这三种运算,矩阵减法就可以通过将要被减去的矩阵乘以-1,然后将两个矩阵相加的方法来实现。类似的,矩阵除法需要找到被除矩阵的逆矩阵,然后将逆矩阵与第二个矩阵相乘。计算矩阵的逆矩阵需要进行一系列的计算,同时并非所有矩阵都有

逆矩阵。尤其是,能进行逆矩阵转换的矩阵必须既是方阵又是非奇异矩阵(也就是非零行列式)。因而,如果 **A** 是方阵且是非奇异矩阵,那么其逆矩阵 $\mathbf{A}^{-1}$ 可以表达为

$$\mathbf{A}^{-1} = \left(\frac{1}{|\mathbf{A}|}\right) \text{adj } \mathbf{A}$$

其中 $|\mathbf{A}|$ 是 **A** 的行列式,adj **A** 是 **A** 的伴随矩阵。

矩阵代数的一个重要的用处是计算回归系数。例如,简单线性回归模型可以用矩阵式表达为 $\mathbf{Y} = \mathbf{X\beta}$,其中 **Y** 是因变量值的列向量,**X** 是自变量值矩阵,**β** 是回归系数矩阵。用矩阵代数进行方程处理,可以求解得到 $\mathbf{\beta} = \mathbf{X}^{-1}\mathbf{Y}$。

——Timothy M. Hagle

(李洋译　高勇校)

参考文献

Cullen, C. G. (1990). *Matrices and linear transformations* (2nd ed.). New York: Dover.

Pettofrezzo, A. J. (1978). *Matrices and transformations.* New York: Dover.

成熟效应(Maturation Effect)

成熟效应是个体随时间推移产生系统性变化的生物过程或心理过程,这一过程与外部事件无关。成熟效应的例子包括随时间而更衰老、更强壮、更英明、更有经验。这些例子涉及长时段的改变。但成熟效应也可能是短暂的,比如变得疲劳、厌倦、高昂、饥饿。

坎贝尔和斯坦利在其经典的《研究中的实验设计和准实验设计》(*Experimental and quasi-experimental designs for research*, Campbell & Stanley, 1963)中,把成熟效应看作能削弱实验内部效度【Internal Validity】的八个竞争性假设之一。内部效度关心的是,能否正确地认定是这一个自变量而非某些外部变量导致了因变量的变化。内部效度是实验的必要条件。

在解释因变量的变化时,能让研究者排除成熟效应和其他一些外部变量最简单的实验设计【Experimental Design】是后测试控制组设计(posttest-only control group design)。按照坎贝尔和斯坦利(Campbell & Stanley, 1963)的论述,这一设计可以图示如下:

$$\begin{matrix} R & X & O \\ R & & O \end{matrix}$$

R 表示将参与者随机分成两组:X 表示处理组,没有 X 表示控制组。控制组和处理组在其他实验实施方面没有任何差别。O 表示因变量的观测值或测量值。事件的顺序是随机分配(R),之后是加以处理或者控制(X),然后是因变量的测量(O)。R 的垂直排列表示参与者在同一时间被分成两个组。同样的,O 的垂直排列表示两个组是在同一时间进行测量的。随机分配【Random Assignment】是这个设计的重要特点。其目的在于确保对自变量进行操作之前两个组的参与者是有可比性的。如果样本规模较大,且每个参与者被分配到不同组的概率相等,那么研究者就有理由认为在两组中已知变量和未知的外部变量的分布是相同的。因此,成熟效应理应在两组中表现相同,由此得以控制。这种简单设计下得到的数据通常可以用独立样本的 t 统计量来分析。

在含有 p 组($p>2$)的实验中,需要对后测试控制组设计进行修正。在修正后的设

计中,$X_1,X_2,\cdots,X_{p-1}$表示各个实验组,X_p表示控制组,如下所示:

$$\begin{matrix} R & X_1 & O \\ R & X_2 & O \\ \vdots & \vdots & \vdots \\ R & X_{p-1} & O \\ R & X_p & O \end{matrix}$$

这种实验得到的数据通常要用完全随机化条件下的方差分析(completely randomized analysis of variance)来进行分析。

——Roger E. Kirk

(李洋译 高勇校)

参考文献

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

最大似然估计(Maximum Likelihood Estimation)

19 世纪末期,卡尔·皮尔逊表明,只要理解了统计分布【Distributions】,研究者就拥有了更好地理解从实验和观察研究中得到结果的有力工具。皮尔逊进一步又指出,只需理解了决定统计分布的少数恒量【Constant】,也就是参数【Parameter】,我们就能理解这个分布的方方面面。只要我们了解了决定分布的参数,我们就不需要"看到"整个分布。在 19 世纪末期,参数统计学这一研究领域诞生了。

这当然是一个很重要的见解。但是难题正在于如何利用从实验和观察研究中得到的经验资料来了解这些分布参数的性质。对这一重要问题,皮尔逊的回答是要进行尽可能多的经验观察,最好对整个总体【Population】进行经验观察,然后计算这些经验分布的各种特征——如均值【Mean】、方差【Variance】、偏度【Skewness】和峰度【Kurtosis】。这些计算出来的经验矩被用来将经验分布映射到一些已知的理论分布上,通常就是所谓的皮尔逊-类型分布之一(Stuart & Ord, 1987)。

大约在同一时期,有个叫罗纳德·费希尔的年轻天才证明了皮尔逊处理估计【Estimation】问题的方法有明显缺陷。据说他比皮尔逊更有才学,因此皮尔逊对他不太大度甚至公开敌视(Salsburg,2001)。费希尔证明,皮尔逊的估计常常是非一致性的(inconsistent)且低效的(inefficient)。费希尔的意思是说,无论收集多少的资料,皮尔逊对概率分布参数的估计通常都不可能在概率上收敛于真正的总体参数(即它们是非一致性的),与其他估计方法相比它们在不同样本之中有较大的波动性(即它们是低效的)。

为了用更合理的估计方法来取代皮尔逊有缺陷的方法,费希尔关注在一个分布参数的不同估计值下能够观测到此数据的概率。更准确地说,他希望找到能够使得观察到所收集数据的概率(通常称之为数据的似然)最大化的参数估计。最大似然估计(maximum likelihood estimates)的优势不仅在于它能将数据似然最大化,而且在于它是一致且有效的,并且后来还发现它在非常广泛的条件下对于诸多模型和分布来说都是渐进无偏的。在 20 世纪早期,最大似然法诞生了。

为了更好地理解这种方法,让我们来考察某个随机变量【Random Variable】再观测 i 的取值。设这个随机变量为 Y_i,研究者观测

到的数据为 y_i。假设我们希望对决定 Y_i 分布的参数进行估计。我们需要知道或者经常要假设 Y_i 分布的形式。可以实现相关矩的一般化，将 Y_i 的已知或假设分布表示为 $f(Y_i;\theta)$，其中 θ 或者是唯一参数或者是决定 Y_i 分布的参数向量。如果 Y_i 是离散的，$f(Y_i;\theta)$ 表示 Y_i 取特定的值 y_i 的概率，也被称为概率函数【Probability Function】。如果 Y_i 是连续的，$f(Y_i;\theta)$ 决定了 y_i 处于某一观测范围内的概率，也通常被称为概率密度函数【Probability Density Function】。

现在让我们假设采用某种方式采集了 N 个数据 $y_1, \cdots, y_N$。只要 $y_1, \cdots, y_N$ 是独立且同分布（independent and identically distributed，即 iid），这些数据是如何收集到的（是观察还是实验）并不重要。也就是说我们需要假设，对于可以生成数据 y_i 和其他一些数据如 $y_{i'}$ 的数据生成机制来说，针对所有的 i 和 i' 都有：(a) 观测 y_i 与观测 $y_{i'}$ 相互独立；(b) 概率（密度）函数 $f(Y_i;\theta)$ 和 $f(Y_{i'};\theta)$ 相同。当然，有一些复杂的方法能够处理这些假设【Assumption】不成立时如何进行最大似然估计。但是现在让我们忽略这些枝节问题。

独立同分布假设成立时，数据的联合概率（密度）函数——即数据似然（data likelihood），或者费希尔称为样本概率——可以表示为个体概率（密度）函数的乘积：

$$L(y;\theta) = \prod_{i=1}^{N} f(y_i;\theta) \tag{1}$$

费希尔的核心观点是，要找到能够使方程（1）（即数据的似然）最大化的 θ 值。由于数值乘积计算有一定难度（小数会下溢，大数会上溢，准确度会因此损失等），人们通常都转而求数据似然的自然对数最大化，或者对数似然最大化，

$$\log\{L(y;\theta)\} = \log\{\prod_{i=1}^{N} f(y_i;\theta)\} = \sum_{i=1}^{N} \log\{f(y_i;\theta)\} \tag{2}$$

这个方程最大化的解与将方程（1）中的函数【Function】最大化的解是相同的。

除了获取数据之外，我们必须设定 $f(y_i;\theta)$ 的形式，以求得使方程（1）和方程（2）最大化的 θ 值。这个取值通常被称作最大似然估计值，常用 $\hat{\theta}$ 来表示。这里有多个分布可以使用。有用的连续分布，包括正态【Normal】、对数正态、Logistic【Logistic】、Beta、指数和 Gamma 等分布；有用离散分布的，包括二项式【Binomial】、多项式【Multinomial】、负二项式【Negative Binomial】和泊松【Poisson】等分布。大多数导读性质的概率和统计教科书都提供了一些研究者可以参考的重点分布（例如 Hogg & Tanis，1988）。伊莱亚森（Eliason，1993）也提供了这方面的案例。选取合适分布的一个重要的考虑因素是要看所观测 y_i 的形式，是否符合候选分布所要求的范围。例如，如果观察到的数据 y_i 都是严格为正的（例如在市场工资中），或者被限定在 0 和 1 之间（例如在概率中），那么选择一个 y_i 可以认为在所有实数（例如服从正态分布）上取值的分布就没有什么意义了。

到目前为止，或者明言，或者暗含，正态分布是社会科学中最经常使用的分布。在一些教学材料中也都假设分布是正态的，然后指出分布均值的最大似然估计量就是样本均值，经常表示为 μ，即

$$\frac{1}{N}\sum_{i=1}^{N} y_i$$

这一结果虽然很重要，但众人皆知，许多材料中都可以找到。糟糕的是，这可能会让人产生这些技术必须进行正态性假设的错误

(incorrect)印象。

所以,为了让分析更有趣,同时表明仅仅需要假设某种(some)分布,在此让我们假设 Y_i 是服从 Gamma 分布的随机变量。Gamma 分布可以采取很多形式,但人们通常将其考虑为类似卡方【Chi-Square】分布的形式,即右偏斜且最左侧位于原点处。它同时也是指数和卡方分布的母体分布(the parent distribution)。如果从社会科学的立场出发,Gamma 分布之所以有趣,可能更重要的在于许多关于奖励及其相关分布(例如,财富、工资、权力等)倾向于服从典型 Gamma 分布。

正如伊莱亚森(Eliason,1993)所提到的,Gamma 分布有两个参数,参数 $\mu = E\{Y_i\}$(即 μ 是 Y_i 的期望均值),且 $\nu = E\{Y_i\}^2/V\{Y_i\}$(即 ν 表示期望值平方与 Y_i 方差的比值),这一分布可以表示为

$$f(y_i;\theta=[\mu,\nu]) = \left(\frac{1}{\Gamma\{\nu\}}\right)\left(\frac{\nu}{\mu}\right)^{\nu} y_i^{(\nu-1)} \exp\left\{\frac{-\nu y_i}{\mu}\right\} \quad (3)$$

为了求出决定分布集中趋势的参数 μ 的最大似然解,我们首先对方程(3)中的 Gamma 密度取自然对数,

$$\log\{f(y_i;\theta=[\mu,\nu])\} = -\log\Gamma\{\nu\} + \nu\log\{\nu/\mu\} + (\nu-1)\log\{y_i\} - (\nu y_i/\mu) \quad (4)$$

进而将这些自然对数相加,得到数据的对数似然,

$$\log\{L(y_i;\theta=[\mu,\nu])\} = -N\log\Gamma\{\nu\} + \nu N\log\{\nu/\mu\} + (\nu-1)\sum_{i=1}^{N}\log\{y_i\} - (\nu/\mu)\sum_{i=1}^{N} y_i \quad (5)$$

下一步我们需要对一阶导数有所了解,因为这里需要计算方程(5)中感兴趣的参数 μ 的一阶导数,以便求出 μ 的最大似然解。尽管计算一阶导数有些困难,但理解一阶导数并不难。也就是说,一阶导数仅仅表示某个函数相对于其他参数的变化率。在这个例子中,某个函数就是方程(5)中的对数似然,而其他参数是参数 μ,即

$$\frac{\partial\log\{L(y_i;\theta=[\mu,\nu])\}}{\partial\mu} = \left(\frac{\nu}{\mu^2}\right)\sum_{i=1}^{N} y_i - \left(\frac{\nu}{\mu}\right)N \quad (6)$$

在微积分中,我们知道当一阶导数(关于 μ)等于 0 时,μ 的相应取值或者最大,或者最小。在这个例子中,我们的确取了最大值——为此我们需要二阶导数的应用知识,这已经超出了本条目的范围。因而,最后一步是将方程(6)中的导数设为 0,并对 μ 进行求解:

$$\left(\frac{\nu}{\mu^2}\right)\sum_{i=1}^{N} y_i - \left(\frac{\nu}{\mu}\right)N = 0 \Rightarrow \hat{\mu} = \sum_{i=1}^{N} y_i/N \quad (7)$$

这就是 μ 的最大似然估计量 $\hat{\mu}$,方程(5)中的 Gamma 对数似然就是样本均值。

重要的是,这些统计技术的适用范围很广。除了 μ,我们还可能对 Gamma 分布中 ν 的最大似然解感兴趣,它是随机变量 Y_i 的方差分量。此外,在社会科学中,我们不仅经常对某个变量的分布参数(例如,那些决定工资、声望和权力等分布的参数)感兴趣,还对其他一些变量(例如,个体、职业、市场和机构的属性)共同影响这些参数的方式感兴趣。这就要求我们将注意力放在问题的另一个层次上——模型设定以及连接变量与感兴趣参数的函数形式。

这里举一个完全简化的例子,我们可能会假定个体属性 X_i(例如性别、种族、年龄、受

教育程度和社会阶层等)嵌入属性为 Z_j(例如,劳动力市场规范的程度、封闭和开放的职业体系、市场集中程度、工会参与率等)的劳动力市场组织,两者相互作用共同影响工资变量的分布。要注意的是 i 表示个体而 j 表示市场。回想一下,皮尔逊在 1800 年代末的研究发现告诉我们,任何变量的分布都完全由其分布的参数所决定。所以,如果我们用有效的方式将这些因素与分布参数联系起来(正如个体和市场属性如何共同影响工资的理论所期待的方式),我们就能够显示这些因素是如何共同对分布产生影响的。

继续讨论 Gamma 分布,我们可以将这些因素与决定工资分布集中趋势的参数 μ 连接起来,而 ν 决定了工资分布的形态。假设 μ 和 ν 在 Gamma 分布中都为正,将这些分布参数与 X_i, Z_j 中的因子连接起来的函数,必须对这些特征有所反映。所以,符合这一条件的最简单的模型,可能会与累加效应(additive effects)建立起对数线性关系。也就是说

$$\mu_{ij} = \exp\{\beta_0 + \beta_1 X_i + \beta_2 Z_j\} \qquad (8)$$

且

$$\nu_{ij} = \exp\{\gamma_0 + \gamma_1 X_i + \gamma_2 Z_j\} \qquad (9)$$

其中 β_1, β_2, γ_1 和 γ_2 都是模型参数的向量,其维度与 X_i 和 Z_j 的维度一致。假定模型设定为方程(8)和方程(9),那么 β 和 γ 的最大似然估计可能包含关于 X_i 和 Z_j 中的个体和市场因素如何影响工资分布的信息。

此外,为了与例子中的嵌入性一致,我们至少还需要一个交互项来影响工资分布。因而,我们仍然需要采用对数线性的连接函数(loglinear link function),但要包含交互项:

$$\mu_{ij} = \exp\{\beta_0 + \beta_1 X_i + \beta_2 Z_j + \beta_3 X_i Z_j\} \qquad (10)$$

$$\nu_{ij} = \exp\{\gamma_0 + \gamma_1 X_i + \gamma_2 Z_j + \gamma_3 X_i Z_j\} \qquad (11)$$

将这些因子用上面的方式与分布参数连接起来,研究人员就能够回答个体和市场属性如何共同影响工资分布这一重要问题了。

尽管从方程(8)到方程(11)的模型增加了新一层的复杂性,但是其基本思想与之前没有建立模型的例子没有太多不同。也就是说面对数据,我们可以(a)为感兴趣的结果变量选择合适的概率(密度)函数,(b)建立函数,将感兴趣的自变量与决定结果变量的概率函数的参数关联起来,以及(c)将对数似然最大化以求得模型参数的最大似然估计值。任何一款允许使用者写出并将任意函数最大化的统计软件,比如 SAS 中的 NLP 过程,都照例可以用来求得最大似然估计值。一旦计算出来,正如上面的例子,这些最大似然估计值就会成为解释个体、市场属性及其交互作用对工资分布效应的基础。

上述分析所展现的核心思想正是最初由费希尔所发展的最大似然法的基础。在当前的计算机软件包中,许多分布和模型参数都是通过最大似然程序和算法估计的,但通常是在后台操作的,因此对像 SPSS 这种点选式统计包的使用者来说看不到这一点。实际上,最大似然蕴含在社会科学研究使用的很多统计模型的估计中都发挥重要作用,包括结构方程模型【Structural Equation Modeling】、Logistic 回归模型【Logistic】、Probit 回归模型【Probit】、Tobit 模型【Tobit】、截尾回归模型【Censored】、事件历史模型【Event History】、对数线性模型【Log-Linear Model】、广义线性模型【Generalized Linear Model】等。因而,理解了最大似然法背后的逻辑,你就能深刻理解这些社会科学家所使用的各种完全不同的模型。

最后,这里有一些有助于更好地理解最大似然估计的文献。伊莱亚森(Eliason,

1993)对这种方法的逻辑和实践进行了易于理解的处理和介绍。金(King,1989)对此也进行了很有益的介绍。吉尔(Gill,2001)的专著介绍了最大似然如何夯实了许多广义线性模型估计的根基。至于更深层次的论述,斯图尔特、奥德和阿诺德(Stuart,Ord,& Arnold,1999)出版了非常有用的专著,对最大似然法与其他估计方法如贝叶斯估计(Bayesian Estimation)技术进行了详细的比较。

——Scott R. Eliason

(李洋译　高勇校)

参考文献

Cramer, J. S. (1986). *Econometric applications of maximum likelihood methods.* Cambridge, UK: Cambridge University Press.

Eliason, S. R. (1993). *Maximum likelihood estimation: Logic and practice.* Newbury Park, CA: Sage.

Gill, J. (2001). *Generalized linear models: A unified approach.* Thousand Oaks, CA: Sage.

Hogg, R. V., & Tanis, E. A. (1988). *Probability and statistical inference.* New York: Macmillan.

King, G. (1989). *Unifying political methodology: The likelihood theory of statistical inference.* New York: Cambridge University Press.

Salsburg, D. (2001). *The lady tasting tea: How statistics revolutionized science in the twentieth century.* New York: Holt.

Stuart, A., & Ord, K. (1987). *Kendall's advanced theory of statistics: Vol. 1. Distributional theory* (5th ed.). New York: Oxford University Press.

Stuart, A., Ord, K., & Arnold, S. (1999). *Kendall's advanced theory of statistics: Vol. 2A. Classical inference & the linear model* (6th ed.). New York: Oxford University Press.

MCA

参见多重分类分析【Multiple Classification Analysis】。

McNemar 变化检验(McNemar Change Test)

参见 McNemar 卡方检验【McNemar's Chi-Square Test】。

McNemar 卡方检验(McNemar's Chi-Square Test)

McNemar 卡方检验也叫作 McNemar 变化检验(McNemar change test),它是与参数检验中的相关样本 t 检验【t-Test】相类似的非参数检验,适用于无法满足参数【Parameter】检验要求的数据。具体而言,这种检验适用于因变量是定类变量(即类别频次)时的情况。这一检验实际上是要看研究对象的类别是否发生了显著变化。

举例来说,让我们考察下弗里多尼亚大学的心理学和社会学系。他们决定让学生在第一学年参加公共社科课程,在这一学年结束时再让学生选择具体学科。有人认为,从改变想法的学生数量上看,社会学系比心理学系从这一过程中获益更多。社会学系建议下一年做个试验,在年末时清点一下改变选择的学生数量,向心理学系

提交一份用他们相信的统计学进行的检验。年初时,系里从250人组成的班级中随机选择了两组学生,每组24位学生,一组学生有强烈意愿学习社会学,另一组愿意学习心理学。年末时,他们比较了年初的专业选择和这些学生下一年的实际专业选择。结果如图1所示。最后有16人选择心理学系,有32人选择社会学系,这是否能说明心理学系总是吃亏呢?或者这仅仅是随机波动造成的差异?

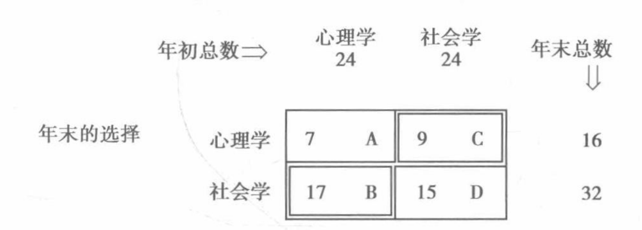


图1 第一年课程前后的学位选择

这可以看作是由针对两组二分变量观测值的前测和后测所组成的实验设计案例,其中每个被试都属于两个备选组中的一个。这里的"处理"似乎是第一年的课程,但也可以说成是了解两个学科以及教师到底如何的共同经历。

对这个测验来说,真正重要的学生是那些改变了想法的学生:C格代表有9名学生从社会学转向心理学,B格代表有17名学生从心理学转向社会学。肯定会有人改变主意,但是如果第一年的经历对两组学生的影响不存在差异,人们会预想从社会学转向心理学和从心理学转向社会学的人应该一样多。换句话说,期望改变可以表达为

$$E = \frac{B + C}{2}$$

很明显,并不是每个样本都可以正好得到这个数,所以我们要用统计检验来考察可接受的变差范围是多少。这里涉及的是类别变量,因此应当采用卡方检验。如果用B和C作为观测频数,将其代入下列卡方比方程中:

$$\chi^2_{\text{ratio}} = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i}$$

方程就会变成

$$\chi^2_{\text{ratio}} = \frac{\left(B - \frac{B + C}{2}\right)^2}{\frac{B + C}{2}} + \frac{\left(C - \frac{B + C}{2}\right)^2}{\frac{B + C}{2}}$$

可以化约为

$$\chi^2_{\text{ratio}} = \frac{(B - C)^2}{B + C}, df = 1$$

(McNemar变化检验)

因此,对于上面的例子来说,

$$\chi^2_{\text{ratio}} = \frac{(9 - 17)^2}{9 + 17} = \frac{64}{26} = 2.46$$

由于$\chi^2(1, 0.05) = 3.84$,结果表明不显著;换句话说,这只是一个偶然事件。可以认为,这一方案最终并不会影响一年后改换专业选择的人数平衡。

——Thomas R. Black
（李洋译　高勇校）

参考文献

Black, T. R. (1999). *Doing quantitative research in the social sciences: An integrated approach to research design, measurement, and statistics.* London: Sage.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Duxbury.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.

均值(Mean)

均值是最常用的描述变量值集中趋势的方法。算术平均数是将变量值加总后除以观测数目——也就是 $\sum x_i/n$，其中 x_i 表示相关变量 x 的第 i 个值，n 是样本量。比如，表 1 显示了 12 个人组成的样本的收入情况。12 个人的总收入是 \$ 3 970，除以 12，则均值为 \$ 331。均值通常就是人们所说的平均数【Average】。x 的均值通常可以表示为 ξ，用来表示 x_i 的最佳估计值，可以将改善后的估计值与之进行比较，如普通最小二乘。均值能在多大程度上表示分布的集中趋势，取决于分布的离散水平，以及是否存偏态分布或"游离值"。比如表 1 虚构的数据显示，个案 F 的收入(944 美元)远远超过排名第二高的收入。如果排除这个偏离较大的个案，均值会降到 275 美元，从整体来看，可以认为此时的均值更能反映典型的收入情况。

表 1　样本 12 人的净收入

人	净收入(美元/周)
A	260
B	241
C	300
D	327
E	289
F	944
G	350
H	195
I	160
J	379
K	214
L	311
总和	3 970
均值	331

正如高尔顿在其关于父亲体重和儿子体重关系的一项著名研究中首次阐明的，随时间推移，极值会越来越接近均值。这也就是趋向均值的回归，也是回归分析的基础。与置信区间【Confidence Interval】结合起来，样本的均值能用来估计给定变量的总体均值。

与算术平均数相比，在社会科学中使用较少的是几何平均数和调和平均数(harmonic mean)。几何平均数是数据值乘积的 n 次根，n 是指个体数量。几何平均数可以表示为

$$(x_1^*, x_2^*, \cdots, x_n)^{1/n}$$

换言之，几何平均数的对数是数据值对数的算术平均数，或

$$\log G = 1/n\left(\sum \log x\right)$$

因此,在表 1 的例子中,G 的对数为 2.47,几何平均数是 296.5 美元。在那些其数据的增长自身与初始值呈几何相关关系时,几何平均数的特性使之更适用于测量均值。比如,几何平均数可以应用在价格(或其他类的)指数上,在价格指数中,价格表示为数列初始值的比值。另一个应用实例是,通过对不同时点上的测量取均值,来对两次测量之间的总体规模进行估计。例如,根据连续两次的 10 年一度人口普查的总体几何平均数,可以在 5 年这个时间中点上生成一个总体估计值。

至于调和平均数,它可以用来测量频次数据或者根据统计值对结果进行分组的数据的集中趋势。调和平均数的倒数是数据值倒数的算术平均数。因此可以表示为

$$1/H = 1/n\sum 1/x \text{ 或 } N \times 1/\sum (1/x)$$

比如,如果在一个样本为 20 个家庭的研究中,7 个家庭有 1 个孩子,8 个家庭有 2 个孩子,5 个家庭有 3 个孩子,家庭规模的调和平均数是 1/(7+4+1.7)/20 = 1.6 个孩子。这低于算术平均数计算出来的 1.9 个孩子。事实上,调和平均数总是低于算术平均数。由于调和平均数计算的倒数性质,当分量频次的关系颠倒过来表达时,它就是算术平均数。因而本例中,在 7 个家庭中,每个孩子有一个家庭;在 8 个家庭中,每个孩子有 0.5 个家庭;在 5 个家庭中,每个孩子有 1/3 个家庭。因此得到每个孩子平均有 0.635 个家庭,反过来说,就是每个家庭有 1.6个孩子。因此,调和平均数也可以用来计算倒数的算术平均数,例如,通过单位美元比或者套数的美元数来计算价格。调和平均数比算术平均数低多少,这取决于离差和均值的大小。

——Lucinda Platt

(李洋译 高勇校)

*也可参见均值【Average】、置信区间【Confidence Interval】、趋中量【Measures of Central Tendency】。

参考文献

Galton, F. (1889). *Natural inheritance*. London: Macmillan.

Stuart, A., & Ord, J. K. (1987). *Kendall's advanced theory of statistics: Vol. 1. Distribution theory* (5th ed.). London: Griffin.

Yule, G. U., & Kendall, M. G. (1950). *An introduction to the theory of statistics* (14th ed.). London: Griffin.

均方误差(Mean Square Error)

在回归【Regression】中,均方误差是总平方和减去回归平方和,然后再根据样本量进行调整(Norusis, 1990, p. B-75)。调整会同时考虑到自由度【Degrees of Freedom】和样本量。均方误差会用来计算测量拟合优度【Goodness-of-Fit】的统计量,包括回归分析中的 R^2 和 F 统计量。

回归中的均方误差和单变量的均方变差(mean squared variation)有如下关系。均方变差是连续变量的离散度的无单位测量(unit-free measure)。均方变差不考虑其偏度。定义均方变差的公式要考虑其样本量以及每个观测值与均值的偏差。

如果我们将均值表达为

$$\bar{X} = \frac{\sum_i X_i}{n}$$

注意这个统计量的自由度是($n-1$),此时均方变差可以表达为(见 SPSS,2001):

$$\text{均方变差} = \frac{\bar{X}}{n-1} = \frac{\sum_i X_i}{n \cdot (n-1)}$$

在多元回归分析中,均方变差是总变差平方和(Total Sum of Squared Variations, TSS)除以样本量 n。但是在计算出含 p 个自变量的回归方程估计值后,均方变差可以分为两个部分:由回归解释的变差和由残差解释的变差。后者就是通常所说的均方误差。定义如下:

$$\text{总平方和} = \text{TSS} = \sum (Y_i - \bar{Y})^2$$

总均方变差是 $\frac{\text{TSS}}{n}$。

$$\text{回归平方和} = \text{RSS} = \sum (\hat{Y} - \bar{Y})^2$$

如下所示,残差误差用 ESS 表示,是 TSS 和 RSS 之差:

$$\text{ESS} = \sum (Y_i - \hat{Y})^2$$

每个平方和项除以各自的自由度就得到了相应的均方项。

回归的均方,或者说均值平方和(Mean Sum of Squares, MSS)是

$$\frac{\text{TSS}}{k}$$

均方误差 MSE 是

$$\frac{\text{ESS}}{n-k-1}$$

如下所示,MSS 与 MSE 的比值则是回归分析的 F 统计量(Hardy,1993,p.22)。

$$F = \frac{\text{MSS}}{\text{MSE}} = \frac{\frac{\text{TSS}}{k}}{\frac{\text{ESS}}{n-k-1}} = \frac{\text{TSS}}{\text{ESS}} \cdot \frac{n-k-1}{k}$$

因此,在多元回归分析中,均方误差对确定 R^2 的显著性上起关键作用。而一般来说,均方的概念在很多统计文本中都有所涉及。

在方差分析【Analysis of Variance, ANOVA】中,总体均方表示因变量的变差,而不同自变量组合所解释的均方都要与这个基准比较来得出。

因此,一般来说,均值平方提供了一种对因变量的离散程度进行无单位测量的测量方法。通过比较总离差平方以及回归平方和,并根据样本量进行调整,我们就可以使用均值平方得到拟合优度的综合指标。根据这些测量(即,R^2、t-检验和 F 检验)进行推论,是进行总体推论显著性判断的标准方法。

在回归分析中,对 R^2 值进行调整很重要;调整后的 R^2【Adjusted R^2】更进一步,根据样本量对检验统计量进行了修正。其公式除了根据给出回归均方和均方误差的自由度进行修正外,还要根据样本量进行调整。

因而,尽管均方误差在回归中作用很大,但在进行推论之前进行进一步的调整还是很有必要的。

——Wendy K. Olsen

(李洋译　高勇校)

参考文献

Hardy, M. (1993). *Regression with dummy variables* (Quantitative Applications in the Social Sciences Series, No. 93). London: Sage.

Norusis, M. (1990). *SPSS/PC + Statistics 4.0*. Chicago: SPSS. SPSS. (2001). *Reliability* (Technical section). Retrieved from www.spss.com/tech/stat/algorithms/11.0/reliability.pdf.

均方(Mean Squares)

一般来讲,均方是对组内个案的平均离差平方进行估计的统计量,同时要根据估计值的自由度进行调整。均方的推论统计必须基于随机抽样【Random Sampling】的假设才能成立。均方也可以用来计算其他的检验统计量。

在回归中,均方误差【Mean Square Error】就是总平方和减去回归平方和,然后再用 n(个案数)和 k(自变量数)加以调整。

均方误差(Mean Square Error, MSE)是

$$\frac{\text{ESS}}{n-k-1}$$

其中的总平方和(Error Sum of Squares, ESS)是误差平方和。

均方误差是对于相对于均值的误差的无单位估计量,同时也考虑了样本规模和自由度的影响。

从两个例子中可以看到均方的两种用法:一是在方差分析【Analysis of Variance】中;二是在回归【Regression】分析中。

在方差分析中,可以得到类似表 1 的表格。

两个均方统计量的比值服从 F 分布,其显著性(在相关的一组假设下)可以从表格或统计包中得到。

表 1 所示的是一个 ANOVA 结果,内容是对市民信心进行的 13 项测量,$n=415$。

在这个表格中,F 统计量的值是 20.080 3 和 0.325 7 的比值(SPSS, 1990, p.868)。如果已被解释的均方与未被解释的均方的比值较高时,就要拒绝不同测量之间不存在显著差异的假设。换句话说,市民信任不仅仅在不同个体之间有所差异,在不同测量之间变化也很大。在这种情况下,方差分析的公式会考虑到样本自由度的计算。

表 1 市民数据的 ANOVA 结果

变差来源	平方和	自由度	均 方	F 统计量
群体之间	669	414	1.617 7	
群体内部	1 859	4 980	0.373 3	
测量之间	241	12	20.080 3	61.66*
残差	1 617	4 968	0.325 7	
总计	2 529	5 394	0.468 8	

来源:SPSS(1990, p.868)。

* 显著性水平<1%。

第二个例子是使用普通最小二乘进行回归估计的结果。在一个对登记工资率的研究中,方程包含受教育年限、全职工作年限、性别和其他变量等31个自变量,结果见表2。

表2 回归均方表

	平方和	自由度	均　方	F统计量	显著性水平
回归	602	31	19.431	129	0.000
残差	549	3 643	0.151		
总计	1 152	3 674			

从表2可以看出,与均方误差相比回归均方较大;调整后的 R^2(adjusted R^2)是52%。在表格中,第一列除以第二列可以得到均方。回归均方除以残差均方可以得到 F 统计量的值,其自由度分别是31 和3 643。统计显著性水平接近于0。

如果运行一个自变量精简后的回归,自由度改变了,但是 F 统计量仍然很显著(见表3);R^2 是38%。如果数据量较小,研究者可能会发现更多自变量的回归方程的 F 统计量会更低,因为自由度损失了。

表3 均方简单回归中 F 统计量的修正

	平方和	自由度	均　方	F统计量	显著性水平
回归	435	10	43	222	0.000
残差	717	3 664	0.196		
总计	1 152	3 674			

这里使用的预测变量是年龄、失业年数、受教育年限、工作时间是否小于30小时、由于照顾家庭而离职的年数、兼职就业年数、隔离程度(某职业中男性总体占比)、全职就业年数、全职工作年数(平方)。在韦尔伯和奥尔森(Walby & Olsen,2002)的著作中可以看到这些变量的详细情况。在这个例子中,自变量精简后的回归 F 统计量更高(为222)。

——Wendy K. Olsen

(李洋译　高勇校)

参考文献

SPSS. (1990). *SPSS reference guide*. Chicago: SPSS.

Walby, S., & Olsen, W. (2002). *The impact of women's position in the labour market on pay and implications for UK productivity*. Cabinet Office Women and Equality Unit, November. Retrieved from www.womenandequalityunit.gov.uk.

量度(Measure)

参见概念化【Conceptualization】、操作化【Operationalization】和量度【Measurement】。

关联量度(Measure of Association)

关联量度是一个统计量,用来考察两个变量共同变动的关系密切程度,或者两个变量的取值模式是否明显不同于每个变量都是随机分布时所观察到的模式。关联量度主要分为两种类型:参数的和非参数的。

对二元关系进行测量之前要注意两个问题:第一,收集数据时所用的抽样具有空间和时间的限制,超出这种时空限制进行统计推论是危险的(Olsen,1993;Skinner,Holt and Smith,1989)。第二,即便证明两个事物之间是强关联,也并不必然说明变量指涉的两个事物之间存在因果关系。

可以测量的关联大致可以分为三种类型:交叉表【Cross-Tabulation】、定序关系、相关【Correlation】。对于包含定类数据(categorical data)的交互表来说,可以使用的检验包括 Phi 系数、Kappa 系数和 Cramer *V* 系数;McNemar 变化检验、费希尔检验和卡方检验【Chi-Square Test】。当一个变量是定序变量,其他变量是分类变量时,可以采用中位数检验【Median Test】,Wilcoxon-Mann-Whitney 检验或者 Kolmogorov-Smirnov 检验【Kolmogorov-Smirnov Test】,以及其他的方法。如果两个变量都是连续性变量,并且符合正态分布,那么皮尔逊相关系数【Pearson's Correlation Coefficient】这一指标可以用来对关联进行从样本到总体的推论。如果两个变量都是定序的,那么可以使用斯皮尔曼秩序相关系数(Spearman's rank-order correlation coefficient)或者 Kendall τ_b。Kendall τ_b 的优势在于可以对多个二元检验进行比较,即使变量是在不同类型的定序尺度上进行测量。

操作化【Operationalization】的决策会影响关联检验的结果。不同的操作化技术要求的测量层次不同。此外,现实的本质也会影响其测量方式(如定类的、定序的、定距的、连续的,以及连续正态分布的测量层次)。

从非参数和参数检验中选择正确的检验方法非常关键。选择检验统计量通常包括三个步骤:对两个变量进行操作化、确定每个变量的测量层次【Level of Measurement】,选择合适的检验方法(参见 Siegel & Castellan,1988)。

下面的例子是一个关于态度的定序测量,但背后基础是连续分布。这个例子体现了配对 *t* 统计量的用法;同时也可以选择针对非正态分布的 Kolmogorov-Smirnov 检验等非参数统计量(表 1)。

表 1　关于印度职业态度调查的成对 t 检验

从事职业	对青年男性从事此项工作的态度的均值	对青年女性从事此项工作的态度的均值	对成对值的成对 t 检验
购买并饲养牲畜	4.2	3.1	-13.8*
用缝纫机缝衣物	2.4	2.8	3.7*

注意:在一对一访谈期间记录受访的倾向性,1=非常不同意;2=不同意;3=中立;4=同意;5=非常同意。

*表示显著性水平为 0.01。

在不同社会背景的个体中,随机抽样出154个人,并询问他们如果可能的话在多大程度上想让一个年轻女性(年轻男性)从事某种类型的工作或培训。他们的答案被编码成一个五分李克特量表,结果是不同的工作有不同的结果。t 检验测量了每一对李克特量表中的关联程序:倾向于年轻男性从事该工作还是倾向于年轻女性从事该工作。配对 t 检验常用在重复测量的情形中。

当地的习俗是年轻男性应该从事牲畜饲养的工作,年轻女性更适合从事衣物缝纫工作。但如同通常的双变量检验一样,还需要进一步确定因果关系的方向,观测到的强关联有可能是虚假相关。

——Wendy K. Olsen

(李洋译　高勇校)

参考文献

Olsen, W. K. (1993). Random samples and repeat surveys in South India. In S. Devereux & J. Hoddinott (Eds.) *Fieldwork in Developing Countries* (pp. 57-72). Boulder, CO: Lynne Rienner.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.

Skinner, C. J., Holt, D., & Smith, T. M. F. (Eds.). (1989). *Analysis of complex surveys.* Chichester, UK: Wiley.

集中趋势量度(Measures of Central Tendency)

集中趋势量度是对数值分布的中心点进行概括的方法。这些方法主要包括均值【Mean】和中位数【Median】。尽管众数【Mode】从严格意义上来说并不是测量集中趋势的方法,但是它也能通过提供分布中出现次数最多的数值来表现分布。因此它也能告诉我们什么是典型的,这也就是我们常常将其与中心点(central point)联系起来的原因。表 1 显示了假想的 7 个家庭的受抚养子女数量。我们可以通过计算算术平均数(取值总和除以观测数)、中位数(或分布的中点)和众数(出现次数最多的值)来对这个分布的中心点进行总结。

表 1　7 个家庭的子女数

家　庭	年龄大于 16 岁的子女数量
Jones	1
Brown	1
Singh	1
Roberts	2
O'Neill	2
Smith	3
Phillips	6

表2显示了在表1的家庭数量下，三种集中趋势量度方法的不同结果。每种方法生成一个对分布中心进行概括的不同数字。在这个例子中，由于变量和观测值范围有限，其中的差异并不显著，但是当取值的范围更广，分布呈现出非对称性时，不同方法得到的值可能会有很大差异。

表2 来自表1的集中趋势量度

测 量	表1的结果
均值	16/7=2.3
中位数	2
众数	1

均值只能用在定距数据上，中位数可以用于定序数据，而众数可以用在定类数据或者定序分类数据上。另一方面，当数据是连续性的或者接近连续性时，众数能提供的信息很少：知道“在一个100人的样本里，两个人的收入是173.56美元，而所有其他样本的收入都不相同”并不能让我们了解这个收入分布。

另一种测量集中趋势的方法很少被用到，就是中列数(midrange)，也就是位于分布两端的最大值和最小值的中间值。在表1的例子中，中列数的值是2.5。与之相关的另一种形式是分布的第25个百分位数和第75个百分位数的中点(即四分位距【Interquartile Range】的中点)。

均值可能根据分布中某一个固定的参照点来计算，通常会设定这个参照点为0。例如，IQ测试将标准智商设为100。一个班级学生的智商平均取值可以根据它们与100的差距大小来确定。最后再用平均差值加上(或者减去)100。这种非零的参考点，通常用于观测频次很大的情况下，此时它能够大大减少计算量。

历史发展

在统计学发展之前的很长一段时间，人们就已经开始使用集中趋势量度的方法了。这是因为人们了解了分布的中间位置之后，也就知道了什么是“正常的”以及自己与它们之间的相对关系。人们通过一个想象的中点来理解他们有多高、多穷、多瘦、多聪明。很长时间以来，算术平均数一直用来检查重复观测值的可靠性程度；但是直到19世纪，统计科学中对集中趋势的测量才开始发展起来。高尔顿第一个用四分位距的中点来作为一种概况性测量方法，在阿道夫·凯特勒关于“均值人(average man)”的著作基础上，他在遗传问题研究中调查了对于均值和中位数的偏离(见平均数【Average】)。弗兰西斯·埃奇沃斯探索了不同的均值和中位数与指数之间的关系，例如价格指数等。集中趋势量度仍然是描述统计学以及理解分布的关键，同时也是理解一些重要的统计概念(如回归)的关键。

集中趋势量度的应用

不同的集中趋势量度方法适用于不同的研究情境或者不同类型的数据。此外，将不同方法结合起来能对于某一分布给出更多信息。如果分布是对称的，均值和中位数相等。但是，如果分布呈现右偏斜(收入分布往往如此)，那么中位数就会低于均值。如果分布为左偏斜(考试结果往往如此)，那么均值就会低于中位数。

表3这个小规模的数据表反映了1999年15个英国家庭在五个变量上的特征，它来自每年的国家家庭调查。对这些变量的位置分布(location)和集中趋势进行总结，所适用的测量方法是不同的。例如，如果考察房间数量，我们可能对均值、中位数或者众数都感兴趣。在这个例子中，房间数目的

均值是5.1,这是个小数,因此并不能表示任何一个家庭的实际房间数目。当我们讲每个家庭的平均子女数是1.7时,这也让人想到把一个孩子分为几份,因此也存在类似的悖论。住房所有权(1 = 完全占有,2=贷款购买,3=承租)是分类变量,它没有任何排序,因此只可以使用众数。相反,尽管当地税级(tax bracket)是定类变量(categorical variable),但是它根据财产的价值进行了排序,因此这里也可以使用中位数。对于家庭收入来说,众数或者中位数(或者同时使用)都是恰当的统计量。读者可以为每个变量计算适合的集中趋势量度。

表3　1999年英国15个家庭的个体属性

区　域	房间数量	住房所有权	家庭收入（欧元/每周）	当地税级
中西部	6	2	202.45	1
中西部	4	1	301	3
中西部	6	3	192.05	2
中西部	6	1	359.02	4
中西部	3	3	209.84	1
中西部	4	3	491.8	1
中西部	5	2	422.7	2
中东部	9	1	434.43	5
中东部	4	3	353	1
中东部	6	2	192	1
中东部	6	2	490.86	3
中东部	5	2	550	4
中东部	3	2	372.88	3
中东部	3	3	303	2
中东部	7	2	693.87	5

最后,我们看到这些家庭都是来自两个地区——中东部和中西部。我们可能想考察不同地区、不同变量的中点是否存在系统差别。这就需要在一个样本的不同组之间或者在不同样本之间进行比较;*t* 检验【*t*-Test】提供了一种进行均值比较并确定两者之间是否存在显著差异的方法,而中位数检验则可以比较不同样本的中位数。

——Lucinda Platt

(李洋译　高勇校)

*也可参见平均数【Average】、四分位距【Interquartile Range】、均值【Mean】、中位数

【Median】、中位数检验【Median Test】、众数【Mode】和 t 检验【t-Test】。

参考文献

Stuart, A., & Ord, J. K. (1987). *Kendall's advanced theory of statistics: Vol. 1. Distribution theory* (5th ed.). London: Griffin.

Yule, G. U., & Kendall, M. G. (1950). *An introduction to the theory of statistics* (14th ed.). London: Griffin.

中位数(Median)

中位数是一个变量分布的中间点,因此分布的一半在中位数以上,另一半在中位数以下。也就是

$$\text{Median} = (n + 1)/2$$

如果个案数目是偶数,中位数必定落在两个中间个案之间,取值为这两个个案的算数平均数。比如,有一组个案

1,2,5,6,8,11,14,17,45

中位数的值是 8。而在这一组个案中

1,5,6,12,18,19

中位数是(6+12)/2=9。

如果数据不是离散数值,而是根据年龄范围或连续变量(比如,用整个阶段的每年的实际年份来表示连续的可能年龄)进行分组,就要用另一种办法来计算中位数。比如,表 1 显示了一个小规模的专业学校中不同年龄段儿童的频次分布。

表 1 按年龄段分小规模专科学校的儿童分布(构造数据)

年龄段(岁)	频次	累积频次
3~5.9	16	16
6~8.9	35	51
9~11.9	62	113
12~14.9	63	176
15~17.9	21	197

这里有 197 名儿童,因此中位数的值是第 99 名儿童的年龄。第 99 名儿童位于 9 到 11.9 岁的年龄组里。事实上,它是这组的第 48 个观测量,因此可以通过计算 3 年的 48/62 是多少,再加上已经过去的 8.9 岁。就是说,此中位数可以这样计算

$$8.9 + [(48/62) \times 3] = 11.2$$

因此,中位数是几种集中趋势量度【Central Tendency】或位置的方法之一,是全面理解分布的一个角度。同时它也是平均数【Average】的一种形式,即便平均数通常被看作算数平均数的另一种说法。尽管中位数不

如均值应用范围广,但在某些方面对理解一个分布确实有优势。主要的优势在于其不易受特异值(outlier)的影响,特异值是一种极端值,是出现的潜在错误或潜在的歪曲个案,它能影响均值但不能影响中位数(见均值【Mean】)。比如,在上面的第一个例子中,45就是一个特异值。此外,当分布的最后一个值未知而导致无法计算均值时,中位数也可以成为一个有价值的平均数。中位数也能在定序数据中使用,这些数据的不同类别有一定的层次顺序,但不同类别之间的距离不同。例如在对教育成就的测量中,第一档是没有学历,第二档是中等学历,第三档是较高学历等。在这种情况下,计算均值没有意义,因为这些值并不表示数量的大小。尽管如此,在这些例子中,中位数则是有意义的。

中位数也可以被视为第50个百分位数,可以说是测量分布的百分位数的一个特殊的例子。百分位数是表明列举出的个案位于该刻度所对应的数值以下的比例,而其余的个案位于该刻度对应的数值以上。因此,第25个百分位数(或者低四分之一位数)表示以此数值为准,25%的个案是小于该值,75%的大于该值。因此,中位数是通过将数值分区来描述分布的一般方法中的一种。一个分布有可以将总体分为100份的99个百分位数,9个可以将总体分成10份的十分位数(deciles)(其中中位数是第5个),以及可以将总体分成4份的3个四分位数(quartiles)(中位数是中间的一个)。

——Lucinda Platt

(李洋译　高勇校)

*也可参见平均数【Average】、集中趋势量度【Measures of Central Tendency】。

参考文献

Agresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Stuart, A., & Ord, J. K. (1987). *Kendall's advanced theory of statistics: Vol. 1. Distribution theory* (5th ed.). London: Griffin.

Yule, G. U., & Kendall, M. G. (1950). *An introduction to the theory of statistics* (14th ed.). London: Griffin.

中位数检验(Median Test)

中位数检验是一种非参数【Nonparametric】检验,用来确认不同样本是否有相同的位置参数,即相同的中位数。也就是说,两个样本来自同一个总体吗?两个样本的中位数的差异是否可以归于抽样变异(sampling variation)?这也可以推广到其他的百分位数,如第25个百分位数(第一个四分位数)或者第75个百分位数(第三个四分位数)。中位数检验起起源于亚历山大·穆德(Alxander Mood)从1950年开始对非参数方法的研究工作(参见Mood & Graybill, 1963)。由于它们关注有序分布(ordered distribution)的秩,它属于样本的线性秩统计量(同样参见Wilcoxon检验【Wilcoxon Test】和Kruskal-Wallis *H*检验【Kruskal-Wallis *H* Test】)。

在中位数检验中,会先计算合并样本(pooled sample)中位数(或其他百分位数),接着计算不同样本中大于这一中位数的个案数量。然后与虚无假设(即两个样本属于同一总体)成立时期望得到的数值比较。所得出的检验统计量服从卡方分布【Chi-Square Distribution】,在某一自由度下的显著性水平也可以计算出来。

例如,让我们考察一个英国 204 位劳动年龄成年人的样本,他们都在 1980 年代中期以前的某个时期经历过失业,其中"失业持续了多少个月"是一个定序变量。我们的问题是男性和女性的失业持续时间的中位数是否相同。在合并样本中,中位数是 2,这表明失业时间在 6 个月和一年之间。表 1 显示了男性和女性中失业持续时间在中位数以上,以及等于或低于中位数的样本数量。

表 1 分性别与中位数相比的失业持续时间

		男性	女性	总和
失业持续时间	>中位数	55	35	90
	≤中位数	44	70	114
总计		99	105	204

在这个检验中,卡方检验【Chi-Square Test】统计量的值为 10.2,自由度为 1,显著水平为 0.001,这表明独立性的假设是不成立的,我们可以拒绝虚无假设。换言之,男性和女性的失业持续时间的中位数是完全不同的。

——Lucinda Platt

(李洋译 高勇校)

参考文献

Hajek, J., & Sidák, Z. (1967). *Theory of rank tests.* New York: Academic Press.

Hollander, M., & Wolfe, D. A. (1973). *Nonparametric statistical methods.* New York: John Wiley.

Mood, A. M., & Graybill, F. A. (1963). *Introduction to the theory of statistics* (2nd ed.). New York: McGraw-Hill.

中介变量(Mediating Variable)

中介变量是在因和果之间起作用的变量。与调节变量【Moderating Variable】类似,中介变量被视为能改变 X 对 Y 的影响的干预因子。中介变量分析就是估计变量 Z 如何影响变量 X 对 Y 的作用,可以用四种方式来估计中介变量作用:路径分析【Path Analysis】、结构方程建模【Structural Equation Modeling】、嵌入线性回归中的代理变量(proxies embedded in linear regression)以及多层次建模【Multilevel Modeling】等。

路径分析对不同的因果进程设立不同的回归方程。递归回归是指会用到其他回归方程中使用的变量的回归(工具变量回归是一个恰当的例子)。在一个回归方程中,变量 X 导致了最终结果 Y,其标准化 β 系数是

$$\beta_i = \mathbf{B}_i \cdot \frac{\mathbf{S}_{\mathbf{X}_i}}{\mathbf{S}_{\mathbf{Y}}}$$

但除此之外,另有一个方程测量了向量 $\mathbf{X}$ 中各元素与 Z 的关联,而 Z 又是与最终结果 Y 相关联的因子。

下面两个例子可以说明 Z 的可能影响。

(1)在对乳腺癌的分析中,社会行为方式如吸烟和决定开始或者停止服用避孕药会影响癌症发展的概率,然而基因和生物因素形成了另一条诱发癌症的路径(Krieger, 1994)。当我们考察病人死于癌症的概率时,这两组影响因子又相互作用,可能抵消相互的影响。因此,形成了一个复杂的路径图,如图 1 所示。

生物学家可能会将社会因子视为"中介变量"。

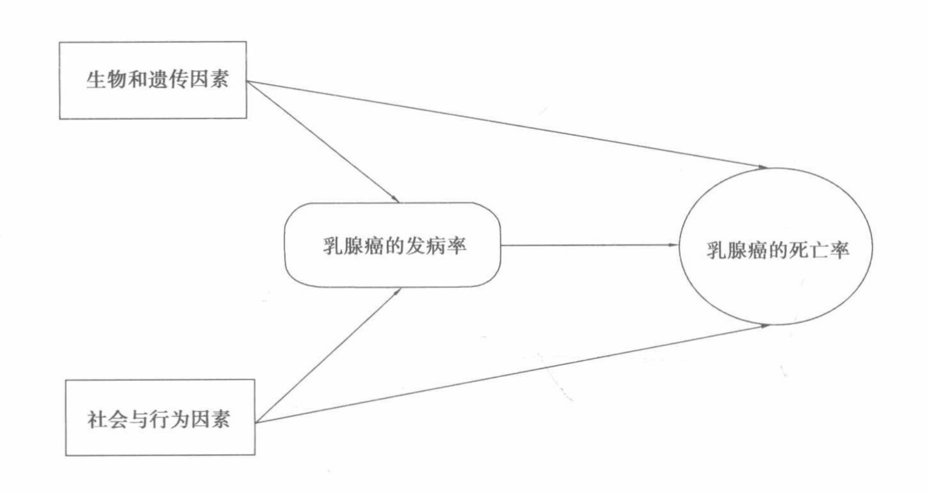


图 1　癌症死亡的原因路径

（2）在分析态度如何与最终结果关联起来时，在社会人口特征对结果的直接影响中，态度可能起到了中介变量的作用。例如，性别并不直接影响外出就餐行为，但是它是一个重要的中介因子。与男性相比，女性可能更希望自己经常外出就餐。因而，性别成为收入和其他财富指标对人们到民族餐厅就餐的总体好奇心产生影响的中介变量（Olsen，1999；Warde，Martens，& Olsen，1999），如图 2 所示。

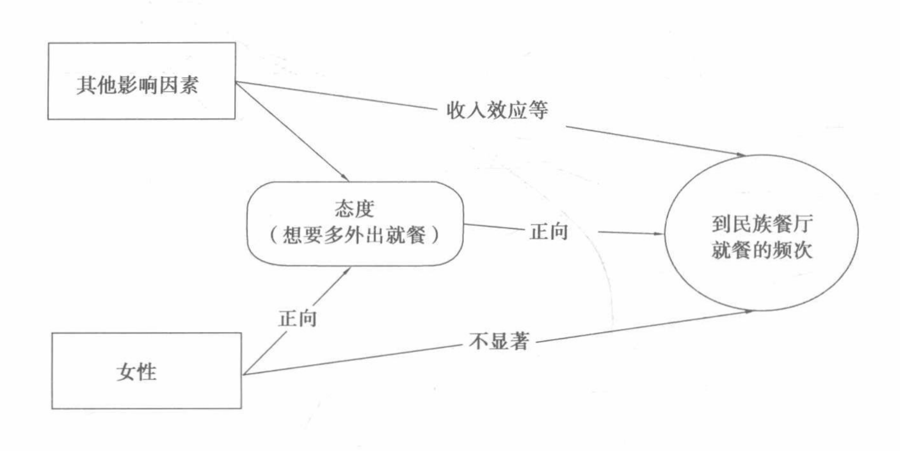


图 2　频繁外出就餐的原因路径

路径分析将误差项视为未被解释的因果过程（Bryman & Cramer，1997，Chap. 10）。在有的经济学分析中，误差项被视为一组未能测量的原因变量，包括一些根本上不可观测的原因变量；在社会学中，误差项被认为是与基本结构模型相对应的个体变差。

最近的进展是，在使用 STATA 软件进行多层次模型估计时，已经能够在模型中对情景效应有所体现，也就是说在这个模型中，个案的分布是条件概率分布（可以对情景效应的或然性质进行建模），并通过多种函数关联起来（就像在标准多层次建模中）（Rabe-

Hesketh, Pickles, & Skrondal, 2001)。因此,随着时间的推移,调节因子的新模型结构可能会出现。

——Wendy K. Olsen

(李洋译 高勇校)

参考文献

Arbuckle, J. (1997). *Amos users' guide version 3.6*. Chicago: Smallwaters Corporation.

Bryman, A., & Cramer, D. (1997). *Quantitative data analysis with SPSS for Windows: A guide for social scientists*. New York: Routledge.

Krieger, N. (1994). Epidemiology and the web of causation: Has anyone seen the spider? *Social Science & Medicine*, *39*(7), 887-903.

Olsen, W. K. (1999). *Path analysis for the study of farming and micro-enterprise* (Bradford Development Paper No. 3). Bradford, UK: University of Bradford, Development and Project Planning Centre.

Rabe-Hesketh, S., Pickles, A., & Skrondal, A. (2001). *GLLAMM manual technical report 2001/01*. London: King's College, University of London, Department of Biostatistics and Computing, Institute of Psychiatry. Retrieved from www. iop. kcl. ac.uk/IoP/Departments/BioComp/programs/ manual.pdf.

Warde, A., Martens, L., & Olsen, W. K. (1999). Consumption and the problem of variety: Cultural omnivorousness, social distinction and dining out. *Sociology*, *30*(1), 105-128.

成员确认与核查(Member Validation and Check)

成员确认与核查也叫作成员核查和被调查者确认,成员确认是一个与定性研究【Qualitative Research】密切相关的过程,研究者将与调查研究相关的材料交给材料提供者进行检查。这个步骤的关键问题是,研究者对在一定社会场景发生的事件的理解与事件成员理解的一致性程度如何。研究者将他或她对研究发现的记录(例如,一个简要报告或访谈手稿)拿出来进行验证,这就是成员核查的最常见的形式。

进行成员确认的主要问题是决定要确认的内容。特别是还有一个问题,即提交上来的材料要在多大程度上包括研究者在面对专业人士写作时进行的各类阐释。即使成员能够证实提交来的材料,故而也能确认研究人员对自己阐释的阐释(the researcher's interpretations of their interpretations),但是不能期望他们从社会科学角度对相关阐释进行演绎。此外,如果成员认为很难理解这些阐释,研究人员也很难知道如何回应才最好。在许多情况下,如果出现了这种情况,研究者需要做的就是修改解释。但是,有时候成员不进行确证,只是因为他们不喜欢研究人员对此的解释。事实上,把解释过的研究材料交给成员,可能就是成员作出防御反应和围绕解释进行争论的开始。

爱默生和波利内尔(Emerson & Pollner, 1988)以其成员确认经验为基础,指出了使用这种方法的另外几个困难:没能阅读报告的所有内容而引发的问题,以及/或者误解所说的内容所引发的问题;很难确定某些评论是赞同还是反对;去询问与作者很亲近的成员,他们不愿意去作出批评,或者无意地提出了一些难以反驳的成员验证问题;成员对于肯定或否定答复的广泛政治含义怀有恐惧心理。布卢尔(Bloor, 1997)同样认为,有时很像研究者期望的那样让成员重视提交的材料,因此漠不关心的态度其实就表明

了一种证实。

就像林肯和古欠(Lincoln & Guba, 1985)所观察到的,虽然成员确认显然不是完美的、没有瑕疵的一个过程,但是它在确立可信度方面起关键作用,同时它还能缓解研究人员对其是否有能力去解释别人的社会世界的焦虑情绪。这种方法能与大多数的定性研究结合起来进行使用。但是,在寻求此种慰藉的过程中,研究者必须对成员确认的局限之处有敏锐的察觉。

——Alan Bryman

(李洋译　高勇校)

* 也可参见受访者验证【Respondent Validation】。

参考文献

Bloor, M. (1997). Addressing social problems through qualitative research. In D. Silverman (Ed.), *Qualitative research: Theory, method and practice* (pp. 221-238). London: Sage.

Emerson, R. M., & Pollner, M. (1988). On the use of members' responses to researchers' accounts. *Human Organization*, *47*, 189-198.

Lincoln, Y. S., & Guba, E. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

成员角色(Membership Roles)

民族志【Ethnography】学者必须承担起与其所研究的世界相适应的社会角色。他们获取信息的途径、他们与被试的关系以及他们所进行的社会学解释在很大程度上受其这些角色特征的影响。早期芝加哥学派的支持者倡导研究立场的多样化,比如完全的观察者(complete observer)、作为参与者的观察者(observer-as-participant)、作为观察者的参与者和完全的参与者(Gold, 1958),其中包括了从最低程度的群体介入到最高程度的群体介入。随后民族志学者被要求客观、超然、保持距离,这些角色就流行起来了。

到1980年代,民族志部分摒弃了其更为客体主义【Objectivism】的倾向而强调主体性(subjectivity)。民族志学者重视那些在深入环境的研究角色。他们寻求能与参加者保持常规的、紧密联系的观察位置。局内人的接纳对他们的角色起关键作用;当在田野研究中拥有成员角色时,个体就不是游离在外的局外人而变成了参与者的身份(Adler & Adler, 1987)。这个身份可以是外围成员角色、活跃成员角色或者完全成员角色。

即使是最边缘的和最不相干的外围角色,仍然需要在群体的社会世界中确认其成员身份。外围成员研究者寻求局内人的视角,接触直观和第一手的经验资料,但并不承担功能性的角色或参与群体的核心活动。在我们对较高层毒品交易者和走私者的研究中,我们融入了其社交共同体和亚文化,被其朋友圈所接受。我们与其成员交往,为其提供住所,照顾他们的子女,形成了长期的朋友关系。然而,关键一点是我们并未参与他们的毒品非法交易活动(参见 Adler, 1985)。

活跃成员边缘化程度较低,在场景设定中也更为中心化,他们参加小组的核心活动,承担一定功能且不仅仅扮演社交或研究角色。除了要具有一定局内人的身份外,他们还与参与者像同事那样互动。在我们对大学运动员的研究中,彼得成为一个篮球队的助理教练。他定期参加球队练习和运作,与球队一同旅行,并且协助球队的常规运作。尽管如此,他的身份和工作能够与其教

练角色分开,教授角色仍然是第一位的(参见 Adler & Adler, 1991)。

完全成员角色需要承担最大的义务。研究者使自己完全沉浸在场景设定中,具备与参与者完全相同的身份,并且有共同的体验和目标。民族志学者会通过选择那些他们已经成为其中一员或可以转换为其中一员的环境,来获得这种角色。尽管这个位置能够进行最深入的理解且虚假程度最低,但这样也更容易导致研究和成员角色需要之间的冲突。在收集青春期前伙伴文化的研究数据时,我们研究了我们的孩子、他们的朋友,以及我们的邻居、朋友、学校和社区。父母的成员角色使我们能自然地进入孩子们的社交网络和"放学"后的活动中。我们使用自己的成员关系来不断进行访谈,并深入了解表面现象背后的知识。尽管如此,有时候,我们内心会有纠结,一方面想要积极地教导这些年轻人,另一方面却需要继续接纳,以便保持开放的沟通这一研究方针(参见 Adler & Adler, 1998)。

或多或少地,成员角色有助于民族志学者在其研究中达到最深的理解层面。

——Patricia A. Adler
Peter Adler
(李洋译　高勇校)

参考文献

Adler, P. A. (1985). *Wheeling and dealing*. New York: Columbia University Press.

Adler, P. A., & Adler, P. (1987). *Membership roles in field research*. Newbury Park, CA: Sage.

Adler, P. A., & Adler, P. (1991). *Backboards and blackboards*. New York: Columbia University Press.

Adler, P. A., & Adler, P. (1998). *Peer power*. New Brunswick, NJ: Rutgers University Press.

Gold, R. L. (1958). Roles in sociological field observations. *Social Forces*, *36*, 217-223.

备忘录,撰写备忘录(Memos, Memoing)

备忘录是研究者的分析记录。写备忘录很花时间,但是它值得做。写备忘录这种方法能够使研究者保持清醒认识,确保研究的诚实性。研究者常常在分析过程中寻求捷径,如在录音转录稿(transcript)的页边做笔记,或者事后写作时根据记忆来填充细节。不幸的是,记忆常常出现错误,分析中闪现的最具深刻见解的"金矿"也会丢失。此外,备忘录有助于提炼分析思路,厘清呈现出来的发现的逻辑。随着备忘录在深度和广度上逐渐加深,研究者可以用备忘录来发现需要进一步调查的领域,以进一步完善研究发现。在团队工作时,备忘录有助于每个人之间相互交流和分享观点(Huberman & Miles, 1994)。在写作阶段,研究者只需要借助备忘录便可以勾勒出写作提纲。

关于如何做备忘录并没有一个硬性的准则。尽管如此,仍然有一些建议能够提高备忘录的质量、强化备忘录的作用。备忘录应当有题目和时间。它们应当是分析性的而非描述性的,也就是说,要关注在分析中出现的概念而非仅仅充斥着原始资料(Dey, 1993)。借助计算机程序,人们可以对概念在原始资料中出现的位置进行标记,并在需要的时候再次找到这些词条(Richards & Richards, 1994)。备忘录应该对资料提出问题,例如,"我从这儿到哪里能够得到这个概念的更多资料,或者我进行下一个观察或者访谈时应该在思考什么?"随着更多的关于这一概念的资料出现,备忘录应该根据它们的属性和维度来对概念进行探索,这些属性和维度包括谁、什么、何时、

在哪儿、怎样、为什么,等等。研究者要随时有一张纸和一支笔,甚至要放在床边。你从来不知道何时灵感会闪现,何时某一块拼图会正好到位。分析人员不应该在进行分析之后很长时间才撰写备忘录。那时不仅当初的想法忘记了,撰写的动力也消散了。分析者不要惧于写一些奇思妙想,即使当时看起来不着边际。如果这些想法没有得到资料的支持,你大可以将之抛在脑后。最后,备忘录可以用来记录自己对研究过程的反应,包括个人的偏见和假设,这些都会不可避免地影响分析的基调和结果。日后写作时,撰写的备忘录将会非常有用。

撰写备忘录能够刺激思维,让你以一种既有创新性又脚踏实地的方式与资料进行互动。在研究项目的初始阶段,备忘录更多会是探索性的、方向性的,而非分析性的。回头看时,这些备忘录常常显得很愚蠢。这是可以预料到的。做研究是逐步推进的,是一种探索行为,你只有在备忘录中才能记录下这一过程。

——Juliet M. Corbin

(李洋译　高勇校)

参考文献

Dey, I. (1993). *Qualitative data analysis*. London: Routledge Kegan Paul.

Huberman, A. M., & Miles, M. B. (1994). Data management and analysis methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 428-444). Thousand Oaks, CA: Sage.

Richards, T. J., & Richards, L. (1994). Using computers in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 445-462). Thousand Oaks, CA: Sage.

元分析(Meta-Analysis)

元分析是为了整合研究发现对多个研究的结果进行的统计分析。吉恩·格拉斯在 1976 年提出这一词语,他认为多个研究的发现积累起来就组成了复杂的数据点,但是如果不进行统计分析,这些数据点并不比单一研究中发现的数据点易于理解。

早在 1976 年以前,在统计教材和文章中已经出现了实施元分析的方法,但这些方法很少得到应用。社会科学中证据的不断扩展和对研究综合的不断增长的需求,才使得元分析得以广泛应用。许多学者希望整合研究文献,从一系列相关研究中得出普遍性结论,但是这一过程却缺乏标准化程序,社会科学研究对这一问题的关注正在激增。在许多研究领域内,已经不可能对每项相关研究都用语言进行描述。一种传统的策略是从数十个或上百个研究中选取一两项研究进行重点关注。这个策略无法对知识的积累进行准确的刻画。首先,这种选择性的关注容易产生确认偏倚(confirmatory bias):一个综合者(synthesist)可能只关注支持其立场的研究。其次,只关注所有研究中一小部分的选择性关注,对于现存检验的整体数量没有考虑或考虑不周。最后,对证据的选择性关注无法对关系强度进行可靠的估计。随着在某一议题上的证据不断累积,研究者对"多少"更感兴趣,而不只是"有没有"。

传统的综合者(synthesists)在考察不同研究结果之间的差异时也会面临各种问题。许多研究在某一个程序方面是相同的,但在其他方面有所不同,综合者可以发现这些研究的结果分布。但是他们很难准确地判断

某一程序上的差异是否影响了研究结果,因为单用某一种方法也会得到不同的结果,所以不同方法得到的结果分布是相互重叠的。

那么,在许多情形下综合论者不得不求助于定量综合技术或者元分析。把定量推论方法应用到研究综合(research synthesis)中,是研究文献不断丰富的必然结果。如果能正确应用统计方法,它们就能增强综合推论的效度。元分析遵循的法则与原初研究(primary research)中进行严格数据分析所需要的推论法则相同,只是其一种扩展而已。

历史发展

在20世纪初期,卡尔·皮尔逊(Pearson,1904)应邀对一种伤寒疫苗进行证据评估。他从11个相关研究中收集数据,计算出每项研究的相关【Correlation】系数。他按结果变量把研究分为两组,然后对两组研究的处理效应测量值取平均数。在平均相关的基础上,皮尔逊得出结论认为其他疫苗更为有效。这是已知最早的定量研究综合。

在皮尔逊开展研究综合之后的3/4个世纪,罗伯特·罗森塔尔和唐纳德·鲁宾(Rosenthal & Rubin,1978)着手开展了一项研究综合,研究人际期望对实验室、班级和车间的行为所产生的效应。他们一共发现了345项与此假设相关的研究。几乎是同时,吉恩·格拉斯和玛丽·史密斯(Glass & Smith,1978)开展了关于班级规模和学业成绩相关性问题的研究回顾。他们发现了基于900 000个学生的725个相关估计值。这一文献展现了833项测试。约翰·赫特、弗兰克·斯密特和龙达·赫特(Hunter, Schmidt, & Hunter,1979)则发现了有关雇用测试对黑人和白人的不同适用性的866项对比研究。

三个研究小组在很大程度上独立地重新发现和创造了针对各自问题的皮尔逊方法。其他人也陆续很快参与到相关研究中。在1980年代,拉里·海奇斯和英格拉姆·奥利金(Hedges & Olkin,1985)用严格的统计证明,将元分析确立为统计科学中独立的专门领域。

对元分析也有一些批评,有些批评一直存在。定量综合的价值被质疑的根据,与对原初数据分析的质疑很相似。也就是说,人们会质疑,使用数字来总结研究文献,会不会让人们印象中误以为的精确性超过了数据本身所能确保的精确性。但是,多数批评并不是针对元分析本身的议题,而是针对不恰当地使用了综合方法,比如缺乏操作细节,而这又被错误地认为是使用定量方法的副产品。

元分析的要素

假设综合者对惊悚广告(fear-arousing advertisement)能否让青少年认为吸烟有害这一问题感兴趣。假设他们找到了八项研究,每项研究都对上述问题进行了考察。其中六项研究结果显示,看不看惊悚广告的两组青少年,其态度没有显著差异;两项研究显示有显著性差异,看过广告的青少年对吸烟更不赞成。其中,一项研究在$p<0.05$时显著,另一项研究在$p<0.02$时显著(都是双尾)。那么综合论者能拒绝广告没有影响这个虚无假设【Null Hypothesis】吗?

综合者可以采用多种方法来回答这个问题。首先,他可以从八个报告中进行筛选,将与其观点相抵触的研究挑出来,认为这些研究在方法上有局限而弃之不管,把其余的支持其观点的研究展示出来,说明这才是真理。这样的研究综合应该给予高度怀疑。它对于解答问题没有任何价值。

计票

其次,综合者可以将收集到的相关研究

结果放入下面三组中的一组，这三组分别是:统计结果显著且表明广告有效的发现，统计结果显著且表明广告让青少年更喜欢吸烟的发现（在本例中没有研究属于这一组），以及统计结果不显著因此无法拒绝惊悚广告没有作用的研究假设的发现。综合论者接着可以声称数量最多的组胜出。在此例子中，虚无假设获胜。

对显著性结果进行计票，这在直观上很有吸引力，所以也经常被使用。但是这个策略过于保守，无法让人接受。问题在于，如果虚无假设成立，概率只会导致 5%的报告显示看过广告会导致人们反感吸烟。因此，如果有 10%或更少（具体数字要取决于研究的总数）的研究是正向且统计显著的，这就表明了广告会引起实际的差异。但是，计票策略需要至少 34%的研究结果是正向且统计显著，才能认定这一假设为真。因此，对统计显著结果进行计票经常会导致人们放弃“处理是有效的”这一假设，而事实上本不该这么做。

在研究综合中，另一种计票的方法是（a）不考虑显著性，只对结果为正和负的数目进行计数，（b）然后应用“符号检验”来确定文献中正向或负向的出现次数是否超出了只依据概率所能出现的次数。计票方法的优势在于能利用全部研究，但缺点在于没有根据研究的样本量来对研究进行加权。这样，一项有 100 位参与者的研究与一项有 1 000 位参与者的研究被赋予了相同的权重。此外，它也没有考虑每个研究中所假设的关系（或者处理效应的估值）强度，因此显示态度发生了正向微小改变的研究与显示态度发生了负向剧烈改变的研究被赋予了相同的权重。但是，对发现中的方向进行计票的方法仍然不失为其他元分析方法的一种有益的补充。

跨研究合并概率

考虑上述方法的缺点，研究综合者接下来要考虑把每个研究结果中的精确概率合并起来。这些方法要求每个研究中的统计检验（a）都是关于同一个假设的，（b）彼此独立，（c）满足原初研究者所做的初始统计假定【Assumption】。

最常用的方法被称为 Z 加总法（adding Z）。在其最简单的形式中，Z 加总法就是对 p 水平下的 Z 得分进行加总，然后再除以推论检验的个数的平方根。在关于惊悚广告的八项研究中，累积 Z 得分是1.69，$p<0.05$（单尾）。因而，虚无假设被拒绝。

所有研究被综合者得到的概率并不相同。与统计显著的结果相比，不显著的结果被得到的可能性更低。这个事实意味着，Z 加总法可能会低估第一类错误（Type Ⅰ error）的概率。罗森塔尔（Rosenthal，1979）将其称为抽屉问题【File Drawer Problem】，并写道“对这个问题较为极端的看法是……各种期刊充斥着 5%的含第一类误差的研究，而实验室里的文件抽屉充斥着 95%的非显著性研究结果（如 $p<0.05$）”（p.638）。问题可能没有严重到这个程度，但是确实存在。

概率合并法克服了在计票中出现的不当权重问题。但是，它本身也有着严重的局限。首先，计票方法过于保守，而概率合并法则极端强劲。事实上，它如此强劲，以至于如果对某个假设或处理效应进行大量的检验就一定会拒斥虚无假设，因此它也就变成了一种无用的方法。此外，合并概率回答的是效应是否存在的问题；它无法回答效应的大小或重要与否的问题。回答“惊悚广告能否让青少年反感吸烟?”常常并不是最重要的问题；重要问题应当是“惊悚广告有多大的效果?”答案或许是零，或者是一个正值或负值。还有一点，研究综合者可能会问“哪些因素会影响广告的效应”。研究者会认识到，对这个问题的回答有助于对吸烟干预的改善和聚焦提出合理建议。面对这些新问题，综合者会去计算平均效应量【Effect

Size】,它非常像皮尔逊相关系数。

估计效应量

科恩(Cohen,1998)将效应量定义为"总体中某一现象存在的程度,或者虚无假设为假的程度"(pp.9-10)。在元分析中,(a)要先对每一项研究的结果计算其效应量(有时是对研究内部进行效应量比较);(b)对不同研究中的效应量进行平均以估计一般效应量(general magnitudes of effect);(c)对不同研究进行比较,以便发现研究结果是否存在差异,以及如果存在差异,哪些研究特征(features of studies)可以对其进行解释。

效应量的估计方法有很多种,但是社会科学文献中流行的主要有两种:第一种是 *d* 指数,它是对两组均值之间的差异的无尺度测量(scale-free measure)。计算一项研究结果的 *d* 指数,需要用两组均值之差除以其平均标准差【Standard Deviation】或者控制组的标准差。第二种测量效应量的方法是 *r* 指数或者相关系数。它测量两个变量间线性相关的程度。在前面有关广告的例子中,*d* 指数是合适的效应量的度量。

当对不同研究的效应量求平均数时,元分析方法需要对效应量进行加权处理。原因是能生成更准确的估计值的研究(也就是样本量较大的研究)应当在合并后的结果中被赋予更高的权重。《研究综合手册》(*The handbook of research synthesis*, Cooper & Hedges,1994)阐述了计算合适权重的方法。此书还介绍了如何估计效应量的置信区间【Confidence Intervals】。检验两个均值间的差异或者相关性大小为零的虚无假设,也可以不用 *Z* 加总法,而用计算出来的效应量的置信区间。

考察效应量的方差

对相关研究进行元分析的另一个优势,是它允许综合者对于为什么研究结果会有差异进行假设检验。继续以惊悚广告为例,综合者可以对所有研究进行分组,然后计算各组研究的平均 *d* 指数。研究者关注的数据特征不同,希望得到的估计值也不同。例如,评估者可能想比较使用了不同的结果变量(吸烟行为的概率或者对吸烟的态度)的研究的估计值是否不同。评估者还可能希望比较不同媒体形式(如印刷媒体和视频媒体)的平均效应量是否不同。评估者还可能想看看广告对不同类型的青少年(如男性和女性)的效应是否有差异。

在计算出不同类型的研究的平均效应量之后,元分析研究者可以对这些因素是否确实与效应量相关进行统计检验。有时这叫作同质性分析(homogeneity analysis)。没有统计学的帮助,评估者只能简单考察不同研究结果的差异,根据研究特征进行随意分类,同时(基于未加说明的推断规则)决定这个特征是否是结果差异的显著预测变量。往好里说,这种方法不准确。往坏里说,这种推论方法不正确。

而元分析提供了一种可以检验研究的不同特征能否解释结果差异的正规方法。效应量并不准确,即使它们都对同一个潜在总体值进行估计,也会在一定程度上有所差异。同质分析使得综合者可以检验抽样误差【Sampling Error】是否就可以解释结果差异,或者研究特征、样本、处理设计或结果测量才能够解释结果差异。如果确实存在差异,与这些差异相对应的平均效应量也就有了新的意义。

文献中有三种检验效应量的差异的统计方法。第一种方法采用标准的推断检验,比如 ANOVA 或者多重回归【Multiple Regression】。效应量是因变量【Dependent Variable】,研究特征是自变量或预测变量。这种方法备受批判,因为其背后的统计假定往往不能成立。

第二种方法将效应量的差异与来自抽样误差的期望差异进行比较。在所有观测

效应要估计的总体值都相同的前提下，这种方法不仅要计算观测到的效应量的差异，还要计算期望的方差。这种方法还要对效应量进行调整，以避免方法上的一些陷阱(methodological artifacts)。

第三种方法是前面已经提到过的同质分析。它能够为研究文献的统计推论提供最完整的指导。与 ANOVA 相似，它将研究按照不同特征进行分组，对各组的平均效应量进行同质检验。实施同质分析相对来说比较简单。其公式在库珀(Cooper,1998)及库珀和海奇斯(Cooper & Hedges,1994)的著作中有详细阐述。

总之，元分析通常包含四组统计量：(a)对结果为正或为负的频次的分析；(b)对多个独立的 p 水平下的推断检验的合并分析；(c)对平均效应量及其置信区间进行估计；(d)检验研究特征是否影响研究结果的同质分析。随着文献量增加，或者作者提供了效应量的估计值的置信区间，人们已经不再需要进行合并概率分析了。除了这些基本成分外，元分析技术还包括许多专门用于整合不同研究的结果的描述和诊断技术。

——Harris M. Cooper

(李洋译　高勇校)

参考文献

Cohen, J. (1988). *Statistical power analysis in the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum.

Cooper, H. (1998). *Synthesizing research: A guide for literature reviews* (3rd ed.). Thousand Oaks, CA: Sage.

Cooper, H., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Research*, *5*, 3-8.

Glass, G. V., & Smith, M. L. (1978). Meta-analysis of research on the relationship of class size and achievement. *Educational Evaluation and Policy Analysis*, *1*, 2-16.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Hunter, J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin*, *86*, 721-735.

Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *British Medical Journal*, *3*, 1243-1246.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, *86*, 638-641.

Rosenthal, R., & Rubin, D. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences*, *3*, 377-415.

元民族志(Meta-Ethnography)

元民族志是一种能够综合多项定性研究的方法，可以基于一组研究来进行诠释(Noblit & Hare, 1988)。与定量的元分析【Meta-Analysis】(Glass,1977)不同，元民族志并不是汇合多项研究的资料，因为这忽略了定性研究因情境不同而不同的本质。元民族志涉及对多项定性研究中的诠释进行综合，其间研究情境也得以综合。这种方法包括进行更强诠释意味的文献回顾、对相似现象的多个解释进行批判性考察、对多个案例研究进行系统比较以形成诠释性的跨案例结论，以及对民族志【Ethnography】研究进行更规范

的综合。在所有这些应用中,元民族志既是归纳性的,又是诠释性的。它从多项研究开始,通过把不同民族志和/或定性叙述用彼此的用语进行“互译(translate)”从而归纳性地得到一种综合,目标是归纳性地对这些研究整体得到一种新的诠释。

元民族志要经过七个阶段才能完成:第一,人们首先要确定一个能够通过综合多项定性研究来解决的学术议题。第二,必须确定现与此学术议题相关的内容,要考察能从研究中得到什么、受众是谁。第三,在对相关研究进行广泛搜索之后,需要仔细阅读相关研究。第四,必须考察各项研究之间的关系,要制作关于关键隐喻、词条、观点和概念及其相互关系的清单。关系可能是相互支持的(研究之间显然存在相似性)、驳斥的(研究之间互相对立)或提供论据的(各个部分的研究推论出一个整体)。第五,对各种研究以一种“某一叙述与另一叙述相似,除了某些方面以外”的类比方式进行“互译”。第六,互译被综合成为一种新的诠释,它能够说明不同研究中的各种诠释。第七,以一些适合于受众的方式(如文本、戏剧、视频、美术等),将综合后的结果表达或展现出来。

例如,在1980年代弗里曼对玛格丽特·米德的萨摩亚民族志的批驳被大众媒体广泛报道。弗里曼(Freeman,1983)尝试对米德(Mead,1928)的研究工作提出质疑,但是诺布利特和黑尔(Noblit & Hare,1988)在其批驳性的综合研究中认为,这些研究分别来自不同的时间阶段、人类学的不同研究范式、不同的性别,甚至不同的萨摩亚地区(一个是美属萨摩亚,一个是西萨摩亚)。简言之,这些研究并不互相矛盾。它们只是论述了萨摩亚文化的不同方面。弗里曼声称他批驳了米德,米德声称驳斥了生物学的决定主义,这些都是错误的。因此,元民族志是一种重新思考各种民族志综合起来的意义的有力工具。

——George W. Noblit

(李洋译　高勇校)

参考文献

Freeman, D. (1983). *Margaret Mead and Samoa.* Cambridge, MA: Harvard University Press.

Glass, G. V. (1977). Integrating findings. *Review of Research in Education*, *5*, 351-379.

Mead, M. (1928). *Coming of age in Samoa.* New York: William Morrow.

Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography.* Newbury Park, CA: Sage.

隐喻(Metaphors)

隐喻是指一种语言表达的修辞格,即一种事物用另一种事物的词汇理解或描述出来。两种事物并不假定真正相同。例如,人们可能用隐喻的方式谈论说一位思想者“照亮了某一议题”。这并不意味着思想者实实在在地点燃蜡烛或者打开手电。在社会科学中,隐喻的使用已经是一个研究主题,它对理论思维的发展贡献巨大。

乔治·莱考夫和马克·约翰逊(Lakoff & Johnson,1980)在其影响深远的《我们赖以生存的隐喻》(*Metaphors We Live By*)一书中提出,隐喻在社会意义和心理学意义上都是很重要的。他们认为文化上的隐喻影响着人们对事物思考的方式。曾经炫目鲜活的隐喻,经年累月的重复之后也可能变成众所周知的习惯用语。众人知晓的隐喻能够提

供一个引导我们理解社会世界的认知框架。例如,政治的竞争性或学术讨论的竞争性,都能通过战争隐喻的这一习惯用法来加以强化,根据不同形势将其描述为受到"攻击""防御"或"击落"。

在社会科学中,隐喻能够成为理论观点的核心组成部分。例如,将社会生活比作剧场的隐喻在社会学历史中影响极大。这个拟剧论的隐喻产生了很多相关词汇如社会演员、社会角色和社会场景。人们可以在欧文·戈夫曼(Goffman,1959)的著作中发现很多新奇的、创新性的隐喻用法,尤其是在他的经典著作《日常生活中的自我表现》(*The Presentation of Self in Everyday Life*)中。

隐喻就像硬币一样,经过循环使用会失去光泽。"社会角色"的概念曾经是一个隐喻式的词语,但是现在已经变成标准的社会学词语了,它的含义可以从字面意义上进行理解。角色理论家(role theorist)假设社会角色是实际存在的,并且可以直接观察得到。较为新近的隐喻——社会世界是可以阅读的文本或者叙述——也经历了相似的历程。正如在罗兰·巴尔特(Barthes,1977)的著作中论述的,那些曾经新奇的、隐喻的内容从重复变为更多地从字面意义上进行理解。方法论实践的发展,无论是识别社会角色或者"阅读文化实践",更加速了从令人激动的隐喻到常规的字面意义的转变过程。

因为有些隐喻本质上就是歧义性的,因此就认为社会科学中的隐喻都始于间接含糊的表达,这种看法太简单化了。心理学家通常使用"意志如机器"这个隐喻。即便是早期的理论家似乎也不确定是否应当将意志理解为"像机器",或者它是否实际上就是机器(Soyland,1994)。但是,随着后来的科学家逐渐接受这样的观点,将隐喻用现成的方法转变成为研究项目,这样隐喻的要素就逐渐让位于平实的要素。实现了这个过程,一个研究领域也得以成功确立,而新一代思想者受到新的隐喻的启发,又会开始攻击现成的研究领域。

——Michael Billig

(李洋译　高勇校)

参考文献

Barthes, R. (1977). *Image music text.* London: Fontana.

Goffman, E. (1959). *The presentation of self in everyday life.* New York: Anchor.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.

Soyland, A. J. (1994). *Psychology as metaphor.* London: Sage.

方法变异(Method Variance)

方法变异指的是测量方法对测量对象的影响。根据经典测试理论,一项研究中所观测到的参与者之间在某个变量上的方差【Variance】,可以归为潜在的真实分【True Score】或研究构念加上随机误差【Random Error】。方法变异是由测评方法所带来的额外方差。坎贝尔和菲斯克(Campbell & Fiske,1959)就举了用斯金纳箱来调节老鼠导致仪器效应的例子,此外还有心理量表导致格式效应的例子。

一种缺少确凿证据但被广泛接受的观点认为,使用同一方法测评不同变量时,方法效应(也叫作单一方法偏差)会导致这些变量之间产生一定程度的虚假共变

(spurious covariation)。例如,很多研究者对把所有变量放在同一份问卷进行调查的研究持怀疑态度,因为题项都在同一份问卷中,使用的共同方法就会在题项之间产生虚假相关。对方法变异的研究质疑了上述假定。例如,斯佩克特(Spector,1987)认为不存在证据能够认为方法效应在组织研究中普遍存在。他的论据之一是,如果方法变异能够产生了虚假相关,为什么问卷研究中还有那么多变量不相关呢?根据威廉森和布朗(Williams & Brown,1994)的分析,在大多数情况下,方法变异的存在将会使得变量之间的相关性变小而非变大。

斯佩克特和布兰尼克(Spector & Brannick,1995)注意到,认为方法(如问卷)会在所有变量中产生方法效应的传统观点过于简单化。他们论述了构念和方法是如何相互作用的,以至于某些用来对特定变量进行测评的方法可能会产生共同的变异。此外,他们提出重要的并非方法自身,而是方法的特征。例如,在对包含个人隐私或威胁性(例如心理调适)问题进行自陈式(self-reports)测量时,社会期望【Social Desirability】众所周知会造成潜在的偏倚,但它并不会对非敏感问题的回答造成偏行。如果采用自陈式测量的方法,构念又较为敏感,一些方差可以归因于社会期望。在处理这些构念时,有必要找到能够控制这些效应的程序。

——Paul E. Spector

(李洋译 高勇校)

参考文献

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.

Spector, P. E. (1987). Method variance as an artifact in selfreported affect and perceptions at work: Myth or significant problem? *Journal of Applied Psychology*, *72*, 438-443.

Spector, P. E., & Brannick, M. T. (1995). The nature and effects of method variance in organizational research. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology: 1995* (pp. 249-274). West Sussex, England: Wiley.

Williams, L. J., & Brown, B. K. (1994). Method variance in organizational behavior and human resources research: Effects on correlations, path coefficients, and hypothesis testing. *Organizational Behavior and Human Decision Processes*, *57*, 185-209.

方法论整体主义(Methodological Holism)

方法论整体主义是方法论个体主义【Methodological Individualism】的反义词,这种方法论认为社会整体要大于个体态度、信念和行为的总合,认为整体经常能决定个体的特征。在哲学和社会科学中,整体主义思想从黑格尔开始就占优势地位,人们通常认为它起源于柏拉图的著作。方法论整体主义(通常缩写为整体主义)在社会科学学科中有许多种形式。尽管马克思、杜威、涂尔干和帕森斯的整体主义观念和重点各有不同,但是都被视为整体主义者。

最著名的整体主义思想家是社会学家涂尔干(Durkheim,1896/1952),他认为社会的影响必须要由社会原因进行解释,对社会进行个体心理的解释是错误的。其观点来自他长期以来在欧洲几个国家进行的自杀

统计研究。尽管自杀率或者上升或者下降，但是在法医看来，归因于个体原因的自杀比例基本保持不变，表明即使像自杀这样的个体行为也受到社会力量的影响。

在人类学和社会学中，整体主义在很大程度上受到达尔文主义的生物学"有机体系"理论的影响。例如，赫伯特·斯宾塞(Dickens，2000，pp.20-24)发现通过结构调整而进行的社会进化与生物学结构进化很相似。帕森斯(Parsons，1968)提出了一个更为复杂的有机系统的观点，他认为与人体器官相似，社会的各组成部分共同发挥作用来保持相互均衡。帕森斯的整体主义是毫不隐讳的功能主义论点，而整体主义和功能主义的批评者坚持认为其功能主义是内含在整体主义观点中的。

波普尔和哈耶克对整体主义进行了政治上的批判(参见方法论个体主义【Methodological Individualism】)，方法论上的批判则认为整体主义的解释是不完整的，因为它没有最终落在人体的信仰、需求和行动上。例如，尽管《凡尔赛条约》和随后的对德国提出的战争赔偿，可能给第二次世界大战的起源提供了一个历史解释的依据，但是如果不考察为这个合约负责的能动者的观点和行动以及其他个体(尤其是希特勒)对此的反应，上述解释也就站不住脚。

涂尔干以及其后很多人相信社会学作为一个学科，尤其要依赖在实体论上等价于物质世界中的物体的社会事实。尽管如此，在最近几年，方法论整体主义和个体主义的说法(以及争论的对象)已经被"结构和能动"所取代。如今只有一些思想家将他们自己称为"整体主义者"或"个体主义者"，尽管两种方法的极端形式都能在逻辑上进行说明，但许多最近的讨论都集中在社会在多大程度上能被个体能动性所解释，或者个体的信仰、需求和行动能在多大程度上被社会所解释。

近几年，吉登斯(Giddens，1984)等社会学家，以及巴斯卡尔(Bhaskar，1998)等哲学家已经在尝试通过提出"结构化理论"在整体主义和个体主义之间架起桥梁，在吉登斯看来，"结构化理论"主要关注社会实践，即结构的生产和被生产(Giddens，1984)。

——Malcolm Williams

(李洋译　高勇校)

参考文献

Bhaskar, R. (1998). *The possibility of naturalism* (2nd ed.). London: Routledge Kegan Paul.

Dickens, P. (2000). *Social Darwinism*. Buckingham, UK: Open University Press.

Durkheim, E. (1952). *Suicide*. London: Routledge Kegan Paul. (Original work published 1896)

Giddens, A. (1984). *The constitution of society*. Cambridge, UK: Polity.

Parsons, T. (1968). *The structure of social action* (3rd ed.). New York: Free Press.

方法论个体主义(Methodological Individualism)

在社会科学中有一个悠久的传统，可以上溯到霍布斯，再从密尔、韦伯延续到更近的霍曼斯和沃特金斯，这一传统认为对社会现象的解释必须回溯到个体能动者(individual agents)的特征和倾向上。在最近几十年中，这种方法论的取向已经多少与个体主义的政治和经济学说(Lukes，1994)合为一体。

大多数方法论个体主义的争论都是相对于方法论整体主义【Methodological Holism】(反过来也相同)进行表述的。个体主义者拒绝后者所主张的在个体背后或者与个体相独立的社会整体(social wholes)的存在,他们并不是说社会整体不存在,而是认为社会世界通过个体根据"场景的逻辑"所采取的行动而存在。也就是说,个体根据特定场景中其他人的期望行为来引导自己的行为,以此生成社会情境(social situations)。复杂的社会情境(如经济系统)被认为来自个体的意向、信念、物质资源和特征的特定安排。在解释给定的经济系统时,对其组成部分的考察(如商业银行的角色)仅仅被视为是一种局部解释,虽然可以成为一种启发性的模型(heuristic models)。完整的解释必须从根本上对真实的或可能是"理想的"(统计上的)的能动者(agents)进行考察。

自密尔以来,自由主义中的个体主义政治经济哲学与个体主义方法论的方法之间就存在联系。在哈耶克和波普尔的著作中,这种联系变得更为明确。波普尔认为"集体的"(整体主义的)解释使我们难以察觉事实的真相,认为所有社会制度的运行都来自个体的态度、决定和行动。他甚至认为从集体主义的实然观点中导出了集体主义的应然观点,最终导致了极权主义(Popper, 1986)。

这种政治争论的是非姑且不论,人们普遍认为这种争论不是从方法论角度进行的推论,方法论则与左翼思想家和右翼思想家都有关系(Phillips, 1976)。但是也有很多从方法论层面提出的反对意见,最重要的一种反对意见是它会导致认为社会特征只能用个体心理来进行描述,然而大多数个体特征(如成为一名银行家、教师或者警察)只能用社会制度(social institutions)来解释。第二种来自唯实论【Realism】的反对意见则关注社会实践或对象的本体论地位。持方法论个体主义的人认为,与有真实属性的客观物体不同,社会现象只是简单的集体主义心理构念。作为回应,整体主义认为物质世界中的许多事物只能通过其观测效应(如重力)来为人知晓,外交政策或者法律主体等社会现象也与之相似,但是它们仍然是真实存在的。

——Malcolm Williams

(李洋译　高勇校)

参考文献

Lukes, S. (1994). Methodological individualism reconsidered. In M. Martin & L. McIntyre (Eds.), *Readings in the philosophy of social science* (pp. 451-459). Cambridge: MIT Press.

Phillips, D. (1976). *Holistic thought in social science.* Stanford, CA: Stanford University Press.

Popper, K. (1986). *The poverty of historicism.* London: ARK.

计量变量(Metric Variable)

计量变量是能进行定量【Quantitativel】测度的变量。数值间的间隔距离是相等的(比如,分数从4~6与分数从6~8的距离相等)。定距【Interval】、定比【Ratio】和连续【Continuous】变量都有计量的测度,因此都应该被视为计量变量处理。

——Michael S. Lewis-Beck

(李洋译　高勇校)

* 也可参见定距【Inteval】。

微观(Micro)

微观是宏观的反义词。微观分析是小范围的,或者仅限个体层次的。宏观分析是大范围的,或者是在集体层次的。

——Michael S. Lewis-Beck
(李洋译　高勇校)

微观模拟(Microsimulation)

与模拟【Simulation】一样,微观模拟的主要目的是使用电脑算法【Algorithm】来生成人的、社会的行为。尽管如此,微观模拟的显著特征在于这个模型【Model】在个人、家族、家庭或者公司等微观【Micro】层次发挥作用。通过对由微观实体组成的、大规模、有代表性的宏观【Macro】总体【Populations】进行模拟,其结论可以应用到整个社会或国家。微观模拟不仅在基础研究【Basic Research】上,而且在辅助政策制订上发挥了重要作用,如帮助美国政府权衡福利改革的不同路径。

——Tim Futing Liao
(李洋译　高勇校)

中层理论(Middle-Range Theory)

中层理论的概念与罗伯特·默顿(Merton 1949,1957,1968)及其《社会理论与社会结构》(*Social Theory and Socical Structure*)有着紧密的联系,它介于日常研究的工作假设和关于社会系统的统一的一般理论之间。默顿批判了社会学对一般理论的探索,如塔尔科特·帕森斯的系统,他相信这种理论太过抽象而无法进行实证检验。尽管如此,他也不认为对变量之间的关系进行经验归纳(empirical generalizations)就足够了。他认为社会学应当致力于发展中层理论,它与可观察资料足够接近以便可以检验,同时又足够抽象以启发系统理论的发展。中层理论基于社会生活中发现的规律性具有因果法则(causal laws)(而非经验描述)的形式。默顿意在通过中层理论来呼应他本人倡导的以经验主义为先导的社会学理论建构。他期待这种程序日后可以发展成一种被经验证实的,进而被夯实为社会学科学基础的社会理论主体。

默顿的信念立基于当时的科学哲学(philosophy of science)和经验导向的社会学。对极端一般性和极端特殊性的智识界限的敏感,可以追溯至柏拉图的《泰阿泰德篇》(*Theaetetus*)以及培根和密尔的著作。在《逻辑体系》(*A System of Logic*)一书中,密尔(Mill,1865/1973)明确支持基于共同性类型(classes of communities)的有限因果理论。在社会学家中,默顿引用了埃米尔·涂尔干的《自杀论》(*Suicide*)作为建构中层理论的经典案例,它基于国别案例构建了有限因果理论。20世纪前五十年,这一理念在研究导向的、关注理论的经验可用性的美国社会学家那里逐渐流行起来。与之相伴,比19世纪晚期的宏大社会学理论(grand theories)更具科学特点的以理论为先导的

研究有了更多发展空间。

以科学调查为构建模板，中层理论一般会提出一个中心论点，从中再得到具体假设，然后检验这些假设。建构中层理论始于某一概念及其关联意象，然后发展出一组相关的理论谜题（theoretical puzzles）。默顿认为，这些谜题的核心是产生各种社会行为、组织和变迁的社会机制的运行。中层理论应当建构社会过程的因果模型（如参照群体理论或社会分化理论）。说明理论成立的前提条件（范围条件）使得这种理论限定在中间层次。这一类型的理论化可以应用于微观-行为层面（如小群体的社会心理学），也可以用于宏观-结构（如社会结构因素之间的相互关系）层次，以及介于两者之间（微观-宏观或宏观-微观关系）的层次。

结构功能主义已经不再时髦（de mode），但是本着中层理论精神的社会学研究不断涌现，其中最著名的来自交换理论家和网络理论家。在过去几年，中层理论原则已然再次成为关注和争论的主题。在建构中层理论是发展社会过程因果模型的关键这一前提和假设的基础上，彼得·赫德斯特伦和理查德·斯韦德贝里（Hedstrom & Swedberg，1998）编写了一本论文集，涵盖了不同学科背景学者的研究，专门对微观-宏观关系进行了更多的关注。这些学者与默顿持类似的观点和想法，主张和提倡进行能够经验验证的、有严格界限的社会学的因果理论建构。

——George Ritzer
Todd Stillman
（李洋译　高勇校）

参考文献

Durkheim, E. (1951). *Suicide.* New York: Free Press. Hedstrom, P., & Swedberg, R. (1998). *Social mechanisms: An analytical approach to social theory.* Cambridge, UK: Cambridge University Press.

Merton, R. (1949). *Social theory and social structure.* New York: Free Press.

Merton, R. (1968). *Social theory and social structure* (3rd ed.). New York: Free Press.

Mill, J. S. (1973). *A system of logic.* Toronto: University of Toronto Press. (Original work published 1865)

Sztompka, P. (1990). *Robert Merton: An intellectual profile.* London: Macmillan.

米尔格拉姆实验（Milgram Experiments）

在米尔格拉姆（Milgram，1963，1974）的服从试验（obedience experiments）中，研究参与者扮演“老师”的角色，他们在相关指令下根据其他参与者所扮演的“学生”的学习功课情况，来对其进行电击，并逐渐提高电击程度。参与者服从实验员指令而施以最高程度（显然会引起极大疼痛感）电击的人数，达到了令人意想不到的高比例。事实上，在实验中并没有实施电击，除了“老师”以外，所有的实验参与者都是米尔格拉姆研究团队的成员。即使这样，不同于任何其他研究的是，在许多方面，这些研究已经成为近40年来讨论社会科学研究中的伦理问题的重要参考。为何这些研究在伦理问题的特定情境中如此引人注目？

（1）米尔格拉姆实验已经成为心理学中最著名的研究，有人甚至认为是最重要的研究。这些实验观察到普通人中存在很高比例会服从指令进行施暴，这被认为揭示了纳粹大屠杀的一个主要起因（Miller，1995）。

研究项目既极其著名又存在极大的伦理争论，这种特殊条件使得这些研究产生了前所未有的影响，成为通常讨论研究伦理问题时一个具有高度指导意义的基准。

（2）米尔格拉姆在第一次出版的著作中（Milgram，1963）对参与者所经历的压力和紧张的气氛进行了生动的描绘：“我观察到一位成熟稳重的商人微笑着自信地进入实验室。在20分钟以内，他就变成一个浑身哆嗦、结巴，并受到强烈精神打击、迅速进入精神崩溃边缘的人”（Milgram，1963，p.377）。从这些报告的上下文中可以看出，米尔格拉姆实验逐渐将参与者暴露在心理危机的可能危险水平上。

（3）黛安娜·鲍姆林德通过在《美国心理学家》（*American Psychologist*，Baumrind，1964）上发表的文章，较早对米尔格拉姆实验进行了伦理批判，这具有里程碑式的意义。鲍姆林德指责米尔格拉姆：（a）没有尽到实验人员应保护参与者的情感福利的义务；（b）实验后的任务报告【Debriefing】不充分；（c）使用有缺陷的方法，得出高度可疑的结果，因而作者认为米尔格拉姆的研究发现与其伦理成本相比不值一提。她认为米尔格拉姆实验的参加者在权威人物面前不仅丢失了信任，也永久性地丢失了自尊。尽管鲍姆林德的批评随后受到了米尔格拉姆及其他人（Miller，1986）的反驳，她的观点仍旧产生了极大的影响，成为日后伦理批评的典范。

（4）米尔格拉姆实验的一个关键特点是，方法中设计进了强制（coercion）内容。那些要求从研究中退出的参与者，被实验者不断激励继续其电击其他参加者的角色。根据当前美国心理学协会（American Psychological Association，APA）的伦理标准，这种强制使得这个服从研究是不合伦理的：

APA伦理法案修正草案6

2001年10月21日

8.02 研究的知情同意【Informed Consent】

在获取如标准3.10“知情同意”中所要求的知情同意时，心理学家要告知参加者下列相关事项：（1）研究的目的，预期持续时间以及过程；（2）他们拒绝参与的权利，以及一旦开始参与也可以从研究中退出的权利。

尽管进行欺骗【Deception】并缺乏真正的知情同意是米尔格拉姆实验的重要特征，但这些因素本身并非米尔格拉姆实验留给后人的遗产的最重要特征。有相当大比例的当代心理学研究（包括APA的伦理守则）赞同对欺骗可以审慎地使用（Epley & Huff，1998）。

总之，米尔格拉姆实验对思考研究中的伦理问题产生了相当深远的影响。围绕这些实验的宝贵的、时常是热烈的争论，几代学生和研究者保持了对研究方法中情感和认知风险因素的敏锐洞察。独立于主要研究者的伦理委员会在开展研究之前对研究进行伦理评估，这种常规操作也是米尔格拉姆实验留给后人的重要遗产。最后，米尔格拉姆实验告诉我们，在开展有重要意义和潜在益处的心理学实验时，不可避免地会存在伦理道德成本。因此，在正确认识和保护研究参与者的脆弱性的同时，研究者探究的权利也要得到尊重。

——Arthur G. Miller

（李洋译　高勇校）

参考文献

Baumrind, D.（1964）. Some thoughts on ethics after reading Milgram's “Behavioral study of obedience.”

American Psychologist, *19*, 421-423.

Epley, N., & Huff, C. (1998). Suspicion, affective response, and educational benefit of deception in psychology research. *Personality and Social Psychology Bulletin*, *24*, 759-768.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, *67*, 371-378.

Milgram, S. (1974). *Obedience to authority: An experimental view*. New York: Harper & Row.

Miller, A. G. (1986). *The obedience experiments: A case study of controversy in social science*. New York: Praeger.

Miller, A. G. (1995). Constructions of the obedience experiments: A focus upon domains of relevance. *Journal of Social Issues*, *51*, 33-53.

缺失数据(Missing Data)

在社会科学中,几乎所有的定量数据【Quantitative Data】都有缺失数据,即某些变量在某些个案上的数据是缺失的。例如,在一个样本中,70%的被调查者的年度收入是可以得到的,但是其余 30%是缺失的。有许多原因可以造成数据缺失。被调查者可能拒绝回答某个问题,或者不小心漏答了这个问题。访谈者可能忘记提这个问题,或者没有记下答案。在纵向研究中,被调查者可能死了,或者在下一次跟踪访谈之前搬家了。如果数据是通过合并不同文件而得到的,还会有匹配错误或者记录丢失的可能。

不管是什么原因,缺失数据通常会给统计分析带来很大的问题。实际上,所有标准统计方法都假设每个个案在所有变量上的信息都是完整的。以前几乎没有指南能够指导研究者如何处理缺失数据。研究者只能选择一些临时性的方法,使用这些方法的理由就是大家都一直这么干。幸运的是,现在有两种处理缺失数据的成熟方法:最大似然法和多重填补法(multiple imputation)。如果满足适当的假设,这些方法有出色的统计特性,现在一般研究者也都能使用这些方法。这个条目首先回顾处理缺失数据的传统方法,然后考察两种较新的方法。要有意义地说明这些方法的特性,首先必须指明这些方法能够成立的假设条件。

假设

为了进行简要说明,我们假定一个数据中只包含两个变量:X 和 Y。我们可以观察所有个案的 X 值,但是 Y 变量上有 20%的数据缺失。如果 Y 变量数据的缺失概率既不取决于 Y,也不取决于 X,就可以说 Y 变量的数据是完全随机缺失(Missing Completely at Random, MCAR)的。用公式表达为 $\Pr(Y \text{缺失} | X, Y) = \Pr(Y \text{缺失})$。许多缺失数据技术只有满足这个相当严格的假设才是有效的。

相比而言,较弱的假设是数据是随机缺失(Missing at Random, MAR)的。这个假设是说 Y 变量上数据缺失的概率可能取决于 X 的值,但在 X 给定时并不取决于 Y 的值。用式子表达就是 $\Pr(Y \text{缺失} | X, Y) = \Pr(Y \text{缺失} | X)$。例如,假设 Y 变量测量收入,X 变量是关于受教育时间的变量。如果收入缺失的概率取决于受教育年限,但是在受教育程度的每个水平上,收入数据的缺失并不取决于收入本身,那么就可以说满足 MAR 假设。显然,如果数据是完全随机缺失的,那么它同样也属于随机缺失。

检验数据是否属于完全随机缺失是完全有可能的。例如,我们可以在特定变量

上,对男性和女性进行比较,来观察两者在特定变量上含缺失值个案的比例是否有差异。只要有差异,就违背了 MCAR 原则。但是,要检验数据是否为随机缺失是不可能的。很明显地,你不能判断高收入人群是否比中等收入人群在收入上存在缺失数据的可能性更小。

如果数据是随机缺失的,而且决定数据缺失机制的参数与模型待估参数完全不同,那么缺失数据被称为是可忽略的(ignorable)。在现实情况中,这种技术前提不太可能被违背。即便违背了,只要数据是随机缺失的,假定缺失数据可忽略的方法仍然表现良好。但是如果将数据缺失机制建模成为参数估计过程的一部分,方法会表现更好。所以在实践上,随机缺失和可忽略性通常可以交替使用。

传统方法

处理缺失数据的最常见的做法——以及实际上所有统计软件默认的方法——是整条删除【Listwise Deletion】或完整数据分析。这种方法的规则非常简单:在进行分析时,如果任何个案在任何一个变量上有缺失数据,都将其删除。因为剩下的个案都有完整的数据,这样在使用各种统计方法时就没有困难了。

整条删除有许多让人青睐的特点。除了很容易实施外,它的优点还在于适用于任何统计方法。此外,如果数据完全随机缺失,整条删除将不会在统计分析中产生任何偏差。这是因为在 MCAR 条件下,有完整数据的样本子集实际上是从初始数据组中抽取的一个简单随机样本【Simple Random Sample】。由于同样的原因,用传统方法估计出的标准误【Standard Errors】也是真正标准误的有效估计值。

整条删除有一个很大的问题,它常常会导致数据中较多个案的遗失。例如,假设研究的目标是对有 10 个变量的多元回归模型进行估计。进一步假设每个变量有 5%的缺失数据,且任意变量数据缺失的概率与数据是否在其他变量上缺失并不相关。那么,平均来看,在整条删除之后一半的个案都会被删除。结果是极大地增加了标准误,降低了统计效力。

其他的传统方法则设计来挽救一些在整条删除中被丢失的数据,但是这些方法通常弊大于利。成对删除【Pairwise Deletion】也称为可用个案分析,是一种包括多元回归【Multiple Regression】和因子分析【Factor Analysis】等各类线性模型通用的简单方法。对于分析中的每对变量来说,可以用在这两个变量上没有缺失数据的所有个案,来估计其相关性(或协方差)。这些相关系数可以组成一个简单的相关矩阵,然后输入线性建模程序中。在 MCAR 的前提下,这一方法产生的估计值是一致的(因而近似于无偏),但是标准误和相关的检验统计量一般都有偏倚的。此外,如果相关矩阵不是正定矩阵,这种方法可能会完全失效。

虚拟变量调整是回归分析中处理自变量缺失数据的一种常用的方法。当变量 X 有缺失数据,构建虚拟(效标)变量 D,使得变量在数据出现时取值为 1,其他为 0;在含有缺失数据的个案上,将 X 的值设定为某个常数(如非缺失个案的均值)。变量 X 和 D 都是模型中的自变量。尽管这种方法看起来可以把所有得到的信息都纳入模型,但即使数据是 MCAR,这种方法也被证明会导致有偏估计。

许多缺失数据的处理方法都可以归为一般的插补法【Imputation】,也就是用一些似乎可信的值来替换缺失数据。所有缺失数据都被插补完之后,就可以在看似没有缺失数据的情况下进行统计分析了。其中一种最简单的方法是无条件均值插补:如果一

个变量有缺失数据,可以用无缺失数据的个案的均值来替代缺失数据。但是众所周知这种方法在许多参数上会产生有偏估计量,特别是方差总是被低估。稍好一些的方法是条件均值插补:如果变量有缺失数据,可以建立起含缺失数据的变量对其他不含缺失数据的变量的线性回归。然后,用估计的回归线来求出缺失数据的预测值。如果数据是 MCAR,那么使用这种方法能够获得近似的回归系数的无偏估计值。

即使人们能够通过参数估计避免偏倚,但所有传统的插补的方法都会低估标准误。原因非常简单。统计分析的标准方法会假设所有数据是真实的。如果一些数据是插补的,插补过程就会引入额外的抽样变异(sampling variability),而此种抽样变异并没有被充分考虑。

最大似然法

大大优于传统方法的一种方法是最大似然法【Maximum Likelihood, ML】。在正确设定的模型下,ML 能够生成一致的(因此,近似于无偏估计)、渐近【Asymptotic】有效的、渐进正态的估计值,即便是在数据缺失的时候(Little & Rubin, 2002)。如果数据的缺失机制满足可忽略性,那么就很容易建构似然函数:只需要对缺失值的传统似然进行简单的加总或求积分。通过传统数值方法可以实现似然的最大化。

虽然 ML 处理缺失数据的一般原则在许多年前已经是众所周知,但是直到最近,用于实践的计算方法才得到广泛使用。在数据缺失的情况下,求 ML 估计值的一种通常的数学方法是 EM 算法。尽管如此,大多数 EM 软件只能在多元正态性(multivariate normality)的假设下求出均值、方差和协方差的估计值。此外,使用这种方法无法求出标准误的估计值。

更有用的方法是把 ML 缺失数据方法和常用的结构方程建模程序结合起来,著名的有 LISREL、AMOS 和 MPLUS。在多元正态性的假设条件下,这些程序能够计算众多线性模型的 ML 估计值,包括线性回归和因子分析。估计标准误可以对缺失数据进行适当的调整。

对于分类数据(categorical data)分析,ML 处理缺失数据的方法已经在 LEM 程序中得到应用。这个软件能对含有缺失数据的对数线性【Log-Linear】模型和潜在类别【Latent Class】模型等进行估计。同样在所有的预测变量都是离散变量的情况下,它也能对 Logistic 回归【Logistic Regression】模型进行估计。

多重填补法

尽管 ML 是一种很好的处理缺失数据的方法,但仍然需要专门的软件进行统计分析。目前,这样的软件只能用于相当有限的一些问题上。多重插补法(multiple imputation, MI)是更为一般化的处理方法,这种方法实际上可以用在任意类型的数据上进行任意类型的分析。此外,这些分析可以在传统的统计软件上完成。

MI 的基本思想很简单。它与为每个缺失数据插补单一数值的方法不同,MI 通过给填补过程引入一个随机成分来生成多个数值。在许多实际应用中,仅仅三个插补数值就可以满足需要,当然越多越好。多重插补法可以形成多个数据组。所以,如果每个缺失数据生成三个填补数值,那么结果就会有三个“完整的”数据组,这些数据组在非缺失数据上的值相同,仅仅在插补数据上有些许差异。

接下来,就要使用传统的软件,对每个完整的数据组进行统计分析。最终,将这些重复分析的结果依据一些简单的规则合并起来。为了进行参数估计,可以取几次重复

测算的估计值的均值。要计算得到标准误，先要对几次重复测算的标准误平方取平均值，然后加上几次重复测算的参数估计的方差（含一个小的修正因子），再开平方。公式为

$$SE(\bar{b}) = \sqrt{\frac{1}{M}\sum_k s_k^2 + \left(1+\frac{1}{M}\right)\left(\frac{1}{M-1}\right)\sum_k (b_k - \bar{b})^2}$$

其中，b_k 是第 k 个重复的参数估计值，s_k 是 b_k 的估计标准误，M 是重复测算的数目。

这种方法能够实现三个目标。首先，在插补过程中适当引入随机变异，以消除常见于确定性插补方法中的参数估计偏倚。其次，多次重复这一过程并取参数估计值的均值，可以得到更为稳定的估计值（抽样变异性更少）。最后，多次重复中的参数估计的变异性使得研究者得到能完全反映插补过程不确定性的、较优的标准误估计值。

与 ML 相比，MI 的统计特性如何呢？如果所有假设成立，MI 生成的估计值是一致且渐进正态的，这一点与 ML 一样。但是，除非进行无穷次的插补（这在实践中是不可能的），其估计值并不是完全有效的（fully efficient）。但是在大多数情况下，只需进行少量插补就足以取得接近完全有效的估计值。

在了解了多重填补法的基本结构后，显然接下来的问题是如何生成填补值。有许多不同的方法可以生成填补值。与 ML 类似，第一步是根据数据选择模型。目前最为常用的模型是多元正态模型，这意味着（a）所有变量都服从正态分布，（b）每个变量都可以表示为其他所有变量的线性函数。这个假设的长处在于所有插补值都建立在线性回归的基础上。由每个线性回归的估计残差构成了一个正态分布【Normal Distributions】，插补值的随机变异正是来源于从这个正态分布中的随机抽取。这些随机抽取值被加到了含缺失数据的每个个案的预测值上。

完成这些工作的细节是更为技术性的工作，不需要在这里深究［谢弗（Schafer，1997）给出了非常详细的方法；艾利森（Allison，2001）提供了技术性稍弱的论述］。幸运的是，目前已经有几种方便获取的、便捷的计算机程序，可以让使用者很容易地生成多重插补值。SAS 最近已经引入了多重插补的程序，以及合并多重插补分析结果的程序。SOLAS 是一种进行 MI 的商业单机程序。免费单机程序主要有 NORM 和 AMELIA。

MI 看起来很复杂，对新手是个挑战。尽管与大多数传统的处理缺失数据的方法相比确实有更多的工作要做，但好的软件可以让 MI 变成一个相对简单的过程。例如，用 SAS 对多元回归进行 MI，人们只需要两条额外的程序代码来进行填补，一条额外的代码在回归过程中，随后的两条代码用来合并结果。

其他方法

在结束之前，需要对当前其他几种仍处在发展中的方法进行简单说明。尽管建立在多元正态模型基础上的 MI 方法是应用最广泛的、最简便的方法，但是其他参数模型可能限定条件更少。例如，可以通过对数线性模型、一般位置尺度模型（general location scale model）（分类变量和定量变量均可）或者是序变广义回归模型（sequential generalized regression model）等来生成插补值。同样还有一些非参数和准参数方法可以用于多重插补。

这里讨论的 MI 和 ML 方法都建立在数据缺失机制满足可忽略性的假设基础上，但

也可以在不满足可忽略性的条件下使用这两种方法,此时 Y 变量缺失数据的概率取决于 Y 值自身。赫克曼处理选择性偏误的方法,就是一个通过 ML 解决上述问题的典范。但是,处理不满足可忽略性假定的缺失数据的方法对于数据缺失机制的模型选择非常敏感。

——Paul D. Allison

(李洋译 高勇校)

参考文献

Allison, P. D. (2001). *Missing data*. Thousand Oaks, CA: Sage.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York: John Wiley.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.

误设(Misspecification)

在经典线性回归【Regression】的诸多假定中,其中之一是研究者设定了“真实的”模型来估计。误设是指这一假定未能满足的各种情况。模型误设对参数估计【Parameter Estimation】和标准误【Standard Error】的影响可能很小也可能很大,这要取决于误设的类型及是否有其他条件可以降低其影响。要检验是否存在各种误设,我们就需要考虑在各种竞争性模型【Model】中进行选择的基础。

误设的类型

模型可能在如下三个方面出现误设:(a)在回归元【Regressor】的选择上;(b)在模型变量的测量上;(c)在随机项或者误差项 u_i 的设定上。首先,由于没有正确选择回归元而造成的估计问题又有如下几种:(a)研究者遗漏了模型中的一个重要变量(参考遗漏变量【Omitted Variable】);(b)包含了一个无关变量;(c)或者误设了 X_i 和 Y 之间关系的函数形式。通常,第一种情况——省略了模型的一个相关变量——会对模型估计造成最严重的后果。具体而言,如果某个变量 Z_i 解释了 Y 的部分变化,那么省略 Z_i 就可能导致参数估计有偏且不一致,除非 Z 完全独立于被包含起来的其他变量。偏倚【Bias】的范围和方向取决于遗漏变量和其他模型变量 X_k 之间的相关性【Correlation】。在模型中增加一个无关变量(即“过度拟合”模型)的危害要相对小一些,因为此时所有的模型参数还是无偏和一致的。但是,此时这些估计值是低效的,因为计算出来的每个参数的标准误中包括了“真实”变量和无关变量 X_k 之间的相关。样本中的相关可能是非零的,所以估计标准误也被不必要地夸大了。误设 Y 和 X 的函数形式可以看作一个过度拟合和拟合不足的混合案例。假设“真实”模型为 $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2^2 + u_i$。我们现在估计一个 Y 是 X_1 和 X_2 的简单线性函数(而非真实模型中的 X_2 的平方)的模型,即 $Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + e_i$,此时 α_0,α_1 和 α_2 的参数估计值和标准误将会由于模型含有无关变量 X_2 却省略了平方项 X_2^2 而受到影响。

除了错误地选择回归元,因变量和自变量的测量误差也可以视为某种形式的模型误设。这是因为测量误差会导致参数估计发生与前面讨论很相似的问题(即估计参数是有偏的、不一致的或者低效的)。最后,正确设定产生随机误差 u_i 的过程也可以看作

模型设定问题。我们通常都设定误差项服从均值为 0、方差为 σ^2 的正态分布[即 $u_i \sim N(0,\sigma^2)$]。现在假定真正的误差是由其他过程产生的(如误差项可能具有异方差性【Heteroskedasticity】或者服从对数正态分布)。这时,我们将再次面对参数估计有偏、不一致或者低效这些问题,具体会出现什么问题则取决于误差项的误设类型。

对误设的检验

优秀的理论【Theory】是我们到现在为止用来避免误设的最重要的工具。但是社会科学中的理论往往薄弱而含糊,因此经验性地检验设定是否正确就很重要了。前面提到的每种设定问题都有相应的一些检验方法。研究者应该先从对初始估计进行残差【Residual】分析开始。如果在回归线上方或下方聚焦的个案有着共同的一些特点,这通常表明存在遗漏变量。残差呈现出一定模式,这表明 X 和 Y 之间存在非线性关系(即函数形式问题),或者表明存在异方差性等其他异常状态。简言之,残差分析是研究人员了解个案及其与设定模型关系的一种重要方法。模型设定是在模型和数据之间反复往返的过程,在理论较为薄弱时更需要如此。

如果人们怀疑某个遗漏变量 Z_i 可能是真实模型的一部分,可以用著名的德宾-瓦特逊的 d 统计量来检验这一假设。按照 Z_i 取值的大小来排列原初的普通最小二乘(OLS)残差,然后用通常的办法计算出 d。如果虚无假设被拒绝,那就表明 Z_i 的确在原初的设定中被错误地遗漏了。对遗漏变量的更一般化的检验方法还有拉姆齐的 RESET 检验和拉格朗日乘数(Lagrange Multiplier,LM)检验。但是 RESET 检验和 LM 检验都不能确定遗漏了哪些变量。这个重要任务只有通过对理论的打磨和对案例的深入理解才能实现,或许残差分析等技术也会有益于此。古雅拉提(Gujarati,2003,pp.518-524)对这些检验方法的具体步骤进行了更为详细的描述。

模型设定和模型选择

我们永远无法确保我们找到了的模型就是“真实的”,所以现代的研究者认为最好将努力的重点放在评估各种竞争模型中哪一个更能充分解释某一现象 Y 上。如果对 Y 有两个竞争模型,其中一个模型中所有的解释变量完全是第二个模型中变量的子集,那么我们可以说第一个模型“嵌套”在第二个模型中。注意此时不同方程中的因变量是完全相同的,因此如调整 R 方【R-Squared】等统计量、F 检验和似然比检验(用于非线性估计量)完全可以用于评估哪个模型在经验上更可取。当一个模型的自变量与第二个模型的自变量只有部分重合或根本不重合时,这两个模型就是“非嵌套的”。例如,国民生产总值(GNP)增长的凯恩斯模型和货币模型就是非嵌套的竞争模型。为了比较非嵌套模型的相对表现,研究人员可能会用到调整后的 R^2 统计量、赤池信息准则等其他的拟合优度测量【Goodness-of-Fit】。克拉克(Clarke,2001,pp.732-735)也介绍了 Cox 检验和 Vuong 检验的相关应用。最后,模型选择自然会引向贝叶斯推论【Bayesian Inference】,特别是当数据不能进行重复抽样时。从贝叶斯视角出发如何进行模型选择,这一主题可以参看利默(Leamer,1978)的著作,克拉克(Clarke,2001)将其应用到了实例中并有所扩展。

——Brian M. Pollins

(李洋译　高勇校)

参考文献

Clarke, K. A. (2001). Testing nonnested models of

international relations: Reevaluating realism. *American Journal of Political Science*, *45*, 724-744.
Gujarati, D. N. (2003). *Basic econometrics* (4th ed.). Boston: McGraw-Hill.
Leamer, E. E. (1978). *Specification searches*. New York: John Wiley.

混合设计(Mixed Design)

混合设计一直以来就是实验法的主流，但近几年来随着一种新的统计方法的发展，混合设计变得更为重要。这种新的统计方法就是混合建模(mixed modeling)，它增强了对混合设计生成的数据进行分析的效度【Validity】。

混合设计这一术语有时被宽泛地指代那些既包括定性【Qualitative】变量也包括定量【Quantitative】变量的研究计划，或者既包括受控变量(manipulated variables)，也包括未受控变量的研究计划。但是，心理学家通常将其定义为既包括被试组间因素(between subjects factors)也包括被试组内因素(within-subjects factors)的正式研究计划，而统计学家一般将其定义为既包括固定因素效应又包括随机因素效应的估计研究设计【Research Design】。

对被试组间因素来说，每个实验单元只可能取此因素的某一个水平。因素的各个水平之间是互斥的，如果是一个平衡设计(balanced design)，各个水平都会有相同数量的实验单元。换句话说，每个因素水平会有相同数量的取此水平的单元。如果涵盖了所有水平而不仅仅是一部分时，此类因素的效应被称为固定效应。

对被试组内因素来说，每个实验单元都可能取此因素的所有水平。如果所取的水平是从更大的总体中抽取的随机样本时，其效应被称为随机效应。区组效应(blocking effects)就可以当作随机效应处理，只要实验区组被视为从要推论的区组总体中抽取的一个样本。

考察这样一个例子：一项实验【Experiment】分性别对100个严重抑郁患者测试药物疗法的效应。研究中的性别有两个水平(男性和女性)，两组人数相同。在此性别就是有固定效应的被试组间因素。另一方面，治疗的药物是从目前所有抗抑郁处方药中随机选择的四种药物(四个水平)。如此抽样就能够估计出抗抑郁药物的总体在治疗抑郁症上的随机(处理)效应。在此药物是被试组内因素。但是要注意，如果实验设计不是要测试全部(抽样得到的)的药物的效应，而是要测试仅仅这四种药物中每种药物的单个效应，此时药物因素效应就应当看作固定效应而非随机效应了。无论哪种情况，如果研究被试是从更大的患者总体中随机选择得到的，那么患者也就成了一个随机效应因素。

有多种多样的混合设计。有许多相当类似，但是每种都需要一些多少有些不同的数据分析方法，具体要取决于处理是如何施加于实验单元的。四种常见的固定设计是重复测量【Repeated Measures】、裂区【Split-Plot】、嵌套【Nested】或分层的，以及随机区组【Randomized-Blocks】。在重复测量设计(也叫作被试组内设计)中，实验单元要被多次测量。这样就收集到了效应的时间序列数据，而每次测量之间的时间间隔是相同的。这类设计是完全随机的，除了处理因素外可以包括也可以不包括被试组间因素。如果有被试组间因素，设计就被称为具有裂区/重复测量结构。

在基本的裂区设计中，实验有大小不同

的两类单元，大单元是由小单元构成的区组。有两类处理：第一类（被试组内因素）施于各个大单元；第二类（被试组间因素）施于每个大区组之内的各个小单元。重复测量设计和裂区设计有相似之处，前者中的处理因素等同于后者中的对大单元［也被称为“全区（whole plot）”］的处理。重复测量设计中的时间因素等同于裂区设计中对较小单元组［也被称为“副区（subplot）”］的处理。但是，对重复测量设计中得到的数据的协方差结构进行分析所需的方法，与分析真正的裂区设计的数据是不同的。

嵌套或分层设计是一个实验因素嵌套入另一个实验因素（嵌套设计也常常称作被试组间设计）。这种设计的一种简单形式，就是实验单元嵌套于各种条件当中。在一个“完全分层（purely hierarchical）”设计中，每组因素有隶属关系（inside-outside relationship）。如果嵌套于其中的是区组时，嵌套设计相当于裂区设计。以区组来嵌套，下属因素组（inside factor groups）是完全嵌入较大的因素组内的。这样就不可能区分下属因素的效应和上下隶属因素之间的交互效应。

在随机化区组设计中，处理被随机性地指派给各个区组中的单元。处理效应通常被认为是固定效应（只考虑那些在实验中存在的处理）。区组效应被认为随机效应，是从对其进行推论的更大的区组总体中抽样得到的。如果一种处理对应于一个区块时，这种设计被称为完全随机区组设计（randomized complete blocks design）。

分析混合设计中得到的数据很麻烦，因为有随机因素。近期发展出来的、目前首选的方法是使用最大似然估计【Maximum Likelihood Estimation】（而非最小二乘【Least Squares】），它不需要假设误差项服从独立同分布。之所以如此，是因为协方差已经在分析中被建模了。甚至存在缺失值【Missing Data】（和不平衡设计），此类混合模型也能够给出合适的误差项。但是，对随机因素进行显著性检验【Significance Testing】的 F 检验要成立，就必须满足球形假定【Sphericity Assumption】。如果（在抽取的总体中）各个处理水平之间的差异的方差相同，就可以认为球形假定（或者环状、Huynh-Feldt 结构、Type-H 协方差结构等类似说法）成立。

——Douglas Madsen

（李洋译　高勇校）

参考文献

Cobb, G. W. (1998). *Introduction to design and analysis of experiments.* New York: Springer-Verlag.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger. R. D. (1996). *SAS system for mixed models.* Cary, NC: SAS Institute, Inc.

Milliken, G. A., & Johnson, D. E. (1992). *Analysis of messy data: Vol. 1. Designed experiments.* London: Chapman & Hall.

混合效应模型（Mixed-Effects Model）

方差分析【Analysis of Variance, ANOVA】模型可以用来研究一个因变量 Y 如何依赖于一个或几个因素。这里将因素定义为分类的自变量，即定类测量层次上的解释变量。因素可以包含固定效应或者随机效应。对于含有固定效应的因素来说，表示每个因素类别效应的参数是固定值（非随机的）。对于含随机效应的因素来说，其类

别效应是随机变量。如果统计推论的目的仅仅是对手头数据存在的那些因素类别得出结论,那么固定效应就足够了。如果因素的分类被视为来自某个总体的随机样本,统计推论的目的是得出能够支撑整个总体的结论,那么就需要引入随机效应。如果模型仅仅含有固定效应【Fixed-Effects Model】或者随机效应【Random-Effects Model】,我们就分别称这个模型是固定效应模型或者随机效应模型。如果模型同时包括固定效应和随机效应,那么这个模型就是混合效应模型。

混合效应 ANOVA 模型的一个典型例子,是对不同组织中不同工种的雇员工作满意度的研究。如果研究者想研究的是有限的几种特定工种,但同时对组织的总体感兴趣,尤其想研究总体中组织间的差异程度,那么就需要选择混合效应模型,其中固定效应用于工种,随机效应用于组织。

正如上述定义以及在随机效应模型的词条中所阐述的,随机效应可以看作某一因素的不同水平的随机主效应。随机交互效应也可以进行类似定义。在上一段提到的例子中,就可以包括一个工种和组织的随机交互效应。

混合效应模型也可以对来自裂区设计【Split-Plot Design】和重复测量设计【Repeated-Measures Design】的数据进行分析。在这些设计中,被试(或者其他单元)在不同的条件下被调查。假设这些条件构成了一个因素 B,被试也被视为一个因素 S。这一设计的混合效应模型就包括了 B 的固定效应、S 的随机效应,以及 $B\times S$ 的交互效应。

线性混合模型

上述模型可以扩展为给一个数值型的解释变量估计一个随机系数。这就扩展为一般线性模型【General Linear Model】,它同时包括固定系数和随机系数,通常被称为线性混合模型(linear mixed model)或者随机系数回归模型(random coefficient regression model)。如果用矩阵的形式,可以表达为

$$Y = \mathbf{X\beta} + \mathbf{ZU} + \mathbf{E}$$

其中,Y 是因变量,$\mathbf{X}$ 是固定效应的设计矩阵,$\boldsymbol{\beta}$ 是固定回归系数的向量,$\mathbf{Z}$ 是随机系数的设计矩阵,$\mathbf{U}$ 是随机系数的向量,$\mathbf{E}$ 是随机残差的向量。标准假设是向量 $\mathbf{U}$ 和 $\mathbf{E}$ 相互独立,且服从均值向量为零、协方差矩阵分别为 $\mathbf{G}$ 和 $\mathbf{R}$ 的多元正态分布。这个假设意味着 Y 服从均值为 $\mathbf{X\beta}$、协方差矩阵为 $\mathbf{ZGZ'+R}$ 的多元正态分布。矩阵 $\mathbf{G}$ 和 $\mathbf{R}$ 的结构还要根据研究设计来进一步加以设定。例如,如果有 K 个随机效应因素,残差是完全随机的,那么模型可以设定为

$$Y = \mathbf{X\beta} + \mathbf{Z}_1\mathbf{U}_1 + \mathbf{Z}_2\mathbf{U}_2 + \cdots + \mathbf{Z}_K\mathbf{U}_K + \mathbf{E}$$

其中,所有向量 $\mathbf{U}_1,\cdots,\mathbf{U}_K$ 以及 $\mathbf{E}$ 都是独立的,它们服从均值向量为零、协方差矩阵分别为 $\mathbf{G}_1,\cdots,\mathbf{G}_K$ 和 $\sigma^2\mathbf{I}$ 的多元正态分布,对矩阵 $\mathbf{G}_1$、$\mathbf{G}_K$ 和标量 σ^2 不再进行进一步限定。这个模型是分层线性模型【Hierarchical Linear Model】的矩阵形式,但并没有假定包含随机效应的因素是分层嵌套的,同时它也是多层次分析【Multilevel Analysis】的基本模型。当 $K=1$ 时,它就是一个二水平的分层线性模型,常用于纵贯数据分析(见 Verbeke & Molenberghs,2000,以及多层次分析的词条)。对于纵贯数据来说,协方差矩阵 $\mathbf{R}$ 也可以设定成特定的非对角形式,以表示自相关关系或者其他类型的时间相依关系。

混合效应模型也可以视为广义线性模型【Generalized Linear Model】的扩展。在此类模型中,广义线性模型(generalized linear models)定义中的线性预测项 $\boldsymbol{\eta} = \mathbf{X\beta}$ 被

Xβ +**ZU** 项所取代。这给出了所谓的广义线性混合模型的一般数学形式,它把分层非线性模型【Hierarchical Nonlinear Models】推广到了随机因素不一定嵌套的情形中去。

——Tom A. B. Snijders

(李洋译 高勇校)

参考文献

Cobb, G. W. (1998). *Introduction to design and analysis of experiments.* New York: Springer-Verlag.

Longford, N. T. (1993). *Random coefficient models.* New York: Oxford University Press.

McCulloch, C. E., & Searle, S. R. (2001). *Generalized, linear and mixed models.* New York: John Wiley.

Neter, J., Kutner, M., Nachtsheim, C., & Wasserman, W. W. (1996). *Applied linear regression models.* New York: Irwin.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data.* New York: Springer.

混合方法研究(Mixed-Method Research)

参见多方法研究【Multimethod Research】。

混合模型(有限混合模型)(Mixture Model, Finite Mixture Model)

有限混合模型是潜在类别分析【Latent Class Analysis】的另一个名称。这些模型假设,所关心的统计模型的参数在被称作潜在类别(latent classes)或者混合成分(mixture components)的不可观测的子集中各不相同。有限混合模型的最重要应用包括聚类、量度、密度估计和随机效应建模【Random-Effects Model】。

——Jeroen K. Vermunt

(李洋译 高勇校)

MLE

参见最大似然估计【Maximum Likelihood Estimation】。

流动表(Mobility Table)

流动表是一个二向列联表【Contingency Table】,其中根据同一变量在两个时点的取值对个体进行分类。它因此区分了"不流动"(即在表格的主对角线上的所有个案)和"流动"(位于表格的其他位置的个案)。在社会科学中流动表很常见(例如研究家庭在两次人口普查之间的地理迁移,或在竞选活动中研究人们在不同候选人之间的选择变化)。在社会学中,代际流动表对被调查者自身的社会阶层和初始社会阶层(通常是父辈的阶层)进行交互分类,代内流动表则对他们在两个时间点上的职业或阶层进行

交互分类。

历史发展

流动表的统计方法是从1950年代开始在社会学中取得实质性的发展的,当时主要是进行指数计算。人们提出了流动率或者Glass指数,即流动表中每个观测频数与统计独立(即所谓"完全流动")假设下相应的期望频次的比率。Yasuda指数和Boudon指数基于对总流动分解为两个分量之和:结构流动或"被迫"流动(边缘分布的不同意味着所有个案不可能都在对角线上)和净流动(用来反映关联的强度)。但是,流动指数的缺点逐渐显现出来。

从1980年代开始,新的范式从两个视角来分析流动表。总不流动率($\sum_i f_{ii}/f_{++}$)、流出百分比(f_{ij}/f_{i+})以及流入百分比(f_{ij}/f_{+j})都是描述绝对(或观测)流动的简单工具。但绝对流动将表格边缘分布的变化与变量间内在关联【Association】模式的变化混杂在一起了。对相对流动的研究,就是用对数线性模型【Log-Linear Model】和优势比【Odds Ratio】来对关联的结构和强度进行分析。

在这些模型中,"拓扑"模型或"层次"模型尤为重要,其基本理念是流动表单元格背后存在着不同的稠密度。在二向乘积(对数线性)模型 $F_{ij} = \alpha\beta_i\gamma_j\delta_{ij}$ 中,如果对所有的 i 和 j 来说都有 $\delta_{ij} = 1$,那么 F_{ij} 就是独立模型。这假定整个表格有着统一的稠密度,也就是说,每条流动轨迹或不流动轨迹都是同样可能的(排除了边缘效应之后)。当 $i = j$ 时 $\delta_{ij} = \delta$,而其他时候 $\delta_{ij} = 1$,这就是被称为同一继承模型(uniform inheritance model)的双密度模型(其自由度比独立模型少1),因为 δ(通常大于1)表明所有类别都具有同一的不流动倾向。假设这种倾向因类别而不同(如果 $i = j$,则 $\delta_{ij} = \delta_i$),这就是准完全流动模型(quasi-perfect mobility model)。豪塞尔(Hauser, 1978)认识到对角线单元格与非对角线单元格也能有相同的密度,因此基于利奥·戈德曼先前的工作,他提出了一般水平模型:如果单元格(i,j)位于第 k 个密度水平,那么 $F_{ij} = \alpha\beta_i\gamma_j\delta_k$。

实例

表1是一个流动表,它对父辈职业和子辈初职进行了交互分类,数据来自1973年美国20~64岁的男性。总的不流动率是39.9% [(1 414+524+…+1 832)/19 912],即子辈五人中有两人的职业与父辈的职业相同。从流出来看,农民的儿子中39.4%仍然是农民(1 832/4 650);从流入来看,农民中80.9%父辈就是农民(1 832/2 265)。这两个数字及其对比同时反映了(并混杂了)交互效应(农民职业的继承倾向)和边缘效应(父辈和子辈中农业这一类别的重要性)。

表1 流动表

父辈职业	子辈职业					
	上层非体力	下层非体力	上层体力	下层体力	农业	总计
上层非体力	1 414	521	302	643	40	2 920
下层非体力	724	524	254	703	48	2 253
上层体力	798	648	856	1 676	108	4 086
下层体力	756	914	771	3 325	237	6 003
农业	409	357	441	1 611	1 832	4 650
总计	4 101	2 964	2 624	7 958	2 265	19 912

表 2 五层次流动模型

父辈职业	子辈职业				
	上层非体力	下层非体力	上层体力	下层体力	农业
上层非体力	B	D	E	E	E
下层非体力	C	D	E	E	E
上层体力	E	E	E	E	E
下层体力	E	E	E	D	D
农业	E	E	E	D	A

为了将交互效应剔除，我们需要建立对数线性模型。完全流动性模型与数据（似然比统计量【Likelihood Ratio】L^2 为6 170.13，自由度为16）相去甚远，但是同一继承模型则有了极大的改善（$L^2=2\ 567.66$，$df=15$，$\delta=2.62$）。因此，存在显著的继承效应。但是继承效应的强度因不同类别而有差异，因为准完全流动模型更优（$L^2=683.34$，$df=11$）。最后，表 2 显示了一个由费瑟尔和豪塞曼（Featherman & Hauser，1978）提出的层次模型，它可以相当完美地再现观测数据（$L_2=66.57$，$df=12$）。设 $\delta_E=1$，估计交互参数是 $\delta_A=30.04$，$\delta_B=4.90$，$\delta_C=2.47$ 且 $\delta_D=1.82$。因此，农业的继承倾向是最为显著的；从下层非体力工作向上移动到上层非体力工作的可能性，要比对应的向下移动的可能性更大（排除了边缘效应之后）。

高级主题

层次模型在变量是定类变量或定序变量时同样可用，但在这种情况下有一些专门的模型：对角线模型、交叉参数模型（crossings parameters models）、同一关联模型（uniform association model）以及行（和/或列）效应模型。假设变量的类别是有序次的，但是我们并不知道这些类别的准确顺序而必须从数据中寻求时，可以采用对数乘积关联模型【Association Model】。

当多个流动表格根据第三个（层）变量（如国家或时间）被聚合在一起时，可以就关联程度是否在这一新的维度上有所变动而进行对数线性建模，这通常会产生许多补充参数。在用对数乘积项对上述变动进行精简建模方面，最近已经有很大进展。例如，这些方法发现法国从 1950 年代到 1990 年代，不管是男性还是女性，其出身阶层和最终阶层之间的相关强度已经稳步降低。

——Louis-André Vallet

（李洋译　高勇校）

参考文献

Featherman, D. L., & Hauser, R. M. (1978). *Opportunity and change*. New York: Academic Press.

Hauser, R. M. (1978). A structural model of the mobility table. *Social Forces*, *56*, 919-953.

Hout, M. (1983). *Mobility tables*. Beverly Hills, CA: Sage.

Vallet, L.-A. (2001). Forty years of social mobility in France. *Revue Française de Sociologie: An Annual English Selection*, *42*(Suppl.), 5-64.

Xie, Y. (1992). The log-multiplicative layer effect model for comparing mobility tables. *American Sociological Review*, 57, 380-395.

众数(Mode)

众数是集中趋势量度【Measures of Central Tendency】中的一种,指的是在一个数值的分布中出现次数最多的一个或多个数值。例如,在 9,8,2,4,9,5,5,2,7,10 的一系列数值中,有两个众数(2 和 9)。

——Alan Bryman

(李洋译　高勇校)

模型(Model)

社会科学家使用模型来表示两个截然不同的(尽管相联系的)概念:经验模型和理论模型。经验模型经常以实证检验的形式出现,通常是统计检验,比如在回归【Regression】中,用来判断一组自变量是否会影响应变量【Dependent Variables】。在这样的经验模型中,自变量【Independent Variables】可以通过理论或观察到的经验规律来确定。

与经验模型一样,理论模型也是对现实的抽象。但是两者的目的并不相同。在经验模型中,分析人员试图弄清楚哪个自变量对感兴趣的因变量有显著影响。然后,用实证检验显示出哪些潜在影响是真实的影响,哪些不是真实的。而在理论模型中,分析人员试图就各种因素如何相互作用产生特定结果提供一个抽象的说法。设计理论模型的目的是理解所谓"资料生成过程(data-generating process)",这个过程是真实世界存在的资料的来源(Morton,2000)。为了理解这个过程,分析人员会对现实世界的结果或事件如何发生提出一种论说或"讲述一个故事",并且以此方式识别出导致这些结果或事件的关键因素和机制。

一些理论模型完全用文字表达,用词汇建构其论证。另一些模型则用数学来建构其论证。依靠数学方法对世界进行预测的模型通常被称为形式模型。这些模型由几个部分组成。首先,有常数或者不可变项(undefined terms),这些是给定的模型特征(如议会的成员)。其次,可变项(defined terms)(如议会成员可以组成委员会)。再次,关于目标和倾向性的假设。例如,分析人员假设委员会由政策关注驱动。最后,会有一个序列(sequence),它详细地表明模型中各个事件的发生次序(例如,在议会中提出法案,委员会决定是否将法案提交议事厅等)。

一旦列出了上述各个部分,分析人员开始尝试对模型求解——也就是识别不同结果发生的条件。有许多模型求解的方法。例如,博弈论【Game Theory】模型通常用分析的方法求解,分析者会找到一个解或均衡。另一方面,动态模型(dynamic models)通常用计算的方法求解,分析者设定变量的初始值,然后通过模拟【Simulation】来解释结果。

与文字或非形式模型相比,形式模型有几个优点和好处。第一,形式模型的假设需要清楚地表述。第二,各个项有更专门的界定。第三,对论证逻辑的详细分析更为清晰和透明。

——Charles R. Shipan

(李洋译　高勇校)

参考文献

Morton, Rebecca B. (2000). *Methods and models: A*

guide to the empirical analysis of formal models in political science. New York: Cambridge University Press.

方差分析模型Ⅰ(Model Ⅰ ANOVA)

方差分析【Analysis of Variance, ANOVA】是考察不同类型的观测在某一变量的均值上是否存在差异的方法。典型的例子是比较实验组和控制组的某种行为。研究的问题是,实验【Experiment】组各个被试的平均行为(用定距【Interval】量表测量)与控制组成员的平均行为是否有显著差异。换句话说,实验处理是否会产生影响?ANOVA是用方差之比来进行显著性【Significance】检验的方法,故而被称为方差分析。如果我们所有有类别的观察而不只是一部分类别,ANOVA就采用固定效应模型【Fixed-Effects Model】。上面的例子中只有实验组和控制组两种类别,所以应当采用固定效应模型。但要注意,固定效应模型并不限于对两个类别进行比较。

ANOVA的典型结果如表1所示。

表1

来源	平方和	自由度	均方	*F* 比率	显著性
组间	10.0	1	10.0	1.21	30
组内	66.0	8	8.3		
合计	76.0	9			

资料来源:Iversen and Norpoth(1987.p.21)。

从"平方和"一栏可以看到对组间和组内变差的测量值,这里要假设组内变差是恒量。为了将这些测量值标准化,我们用每个平方和除以各自的自由度。这样我们就得到了两个方差【Variance】(或称为均方【Mean Square】)的估计值。倘若组间方差不大于组内方差,两个量的比率(*F* 比率【*F* Ratio】)应该是1.0。在这种情况下两组之间的差异完全是由于随机概率(random chance)造成的,而不能归因于任何组的特征。在上面的案例中,*F* 比率的值仅有1.21,在0.05的水平上并不显著。因而我们认为实验处理没有影响。

固定效应模型并不只限于检验一个解释变量(单维【One-Way】ANOVA)的效应。它也可以适用于两个解释变量(二维【Two-Way】ANOVA)的情况,此时其中一个是行变量,另一个是列变量。但是,最好把所有行列组合(单元格)的观测数量设计为相等。这当然只有在调查者可以实施控制的实验中才有可能。对于问卷调查或者其他观测数据来说,这一类别组合里的数量取决于实际观测情况。在这种情况中,是无法估计出每个解释因子能够解释的独特方差的。那样也就限制了二维ANOVA对实验设计的作用。

除了对每个解释变量的效应(主效应【Main Effect】)进行检验外,二维ANOVA也能对变量之间的交互【Interaction】效应进行检验。如果两个变量(X_1 和 X_2)对因变量(Y)的主效应不是可加性的(additive),称这两个变量对因变量有交互效应。换句话说,此时即便不考虑随机误差,将两个主效应相加也不能预测因变量。同时被置于两个实验条件下会对研究行为产生的效应,要大于(或小于)对两个单独效应的加总。如果没有周密的理论来说明交互效应的方向,这种类别变量的交互效应通常都很难解释。

——Helmut Norpoth

(李洋译　高勇校)

*也可参见固定效应模型【Fixed-Effects Model】。

参考文献

Hays, W. L. (1994). *Statistics* (5th ed.). New York: Harcourt Brace.

Iversen, G. R., & Norpoth, H. (1987). *Analysis of variance* (2nd ed., Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-001). Newbury Park, CA: Sage.

Scheffe, H. (1959). *The analysis of variance*. New York: John Wiley.

方差分析模型Ⅱ(Model Ⅱ ANOVA)

参见随机效应模型【Random-Effects Model】。

方差分析模型Ⅲ(Model Ⅲ ANOVA)

参见混合效应模型【Mixed-Effects Model】。

建模(Modeling)

建模是创造一个"对某一系统或某一现象的简化表达"的过程,它"要用某种假设【Hypotheses】来描述系统或解释现象"(Random House,1997)。通过建模过程,分析者试图确证或否定他们对于所研究的系统或现象提出的假设,往往要用定量数据【Quantitative Data】来进行假设检验以便完成这一任务。分析者希望借以得到结论的过程是重要的、科学可靠的、用同一模型【Model】和数据可复制的。得到的结论应当可以有助于分析者和读者解释过去的事件,预测未来的行为,提出能更好地解释和预测社会行为的崭新的概念框架。

历史

模型建构在社会科学研究中的流行始于20世纪中期社会科学中发生的行为主义革命(behavioralist revolution)。当时社会科学家(尤其是经济学家)受到自然科学(生物学、化学和物理学)在数学和逻辑上日渐成熟的启发,开始非常想像自然科学家一样找到可测量变量之间的"普遍"关系。寻求普遍关系背后的目的,是为了为人类行为建立类似于法则的论述。人类行为的变异性,特别是在社会科学领域内,应通过科学的使用独立于所研究行为的控制变量来进行解释。布莱洛克(Blalock,1969)认为,归纳性建模是精确的理论论述所必需的,如有必要,还需要基于假设检验的结果对最初提出的理论进行判断。在过去数年间,利用逻辑论演来演绎出形式模型(formal models)的研究越来越流行,在社会科学文献中获得了与归纳模型同等的地位。

如同自然科学家用加速度乘以质量就可得到力的大小一样(用公式表示为$f=m\times a$),行为主义的社会科学家也能够较为精确地知道多大的自变量【Independent Variable】就能够对因变量【Dependent

Variable】(即所研究的人类行为)产生重要影响。政治学和政治社会学中很多人选择去研究选举行为,就是因为社会科学家越来越理解了模型的力量,在模型中加入几个自变量后,就能够了解选举期间个体投票选择的差异原因。模型中的每个自变量都解释了方差的某一部分。

模型建构

社会科学家建构模型以更好地理解现实环境,尤其是通过建模来解决因果问题。如何解释市长候选人布朗所得到的票数?如何解释世界范围内不同国家政体类型的变化?什么因素最能解释美国各州在社会福利开支上的差异?

社会科学模型创造出了一个虚构和抽象的社会环境来表现其最重要的特征。例如,社会科学家想解释选民的投票选择,就要确定选民支持某一候选人、反对另一个候选人的最重要因素是什么,是党派认同、意识形态还是对经济状况的态度。然后通过测量这些重要(尽管无法穷尽)的变量,分析者关注于理解什么才是现实环境的本质。这样就可以更好地理解所研究的现象(即解释了投票选择中的变异性)。

模型类型

模型可以是概念模型,也可以是数学模型。社会科学家的目标是把概念模型转化为数学模型。两者的差异是,概念模型是用文字形式来表达不同的理论,而数学模型要对所有现象进行数值测量得到变量来实现"操作化(operational)"。数值测量可以是直接或间接的。直接测量的一个例子是对特定候选人的选举投票,表示这个候选人获得支持以赢得选任职位的情况。间接测量的一个例子是民主指数,这是用多个测量成分组合起来以代表"民主"概念的。模型进一步具体设定了变量相互影响导致某种行为的路径。因而,数学模型之所以是可以"检验的",就在于人们可通过统计技术来对其进行估计,并得出结论来验证关于现实世界的假设。

模型建构和评估

例如,为了更好地理解不同国家民主政体的产生,研究人员用下列方法创建一个概念模型(conceptual model):

$$民主 = f(经济、文化)$$

研究人员可以进一步改善这一模型,他构建了对上述概念操作化的如下变量:

$$民主=f(经济发展、收入分配、宗教实践、种族构成)$$

在这个模型中,经济发展和收入分配(在全国总人口中的)代表"经济",而整个国家的宗教实践(如全体人口中新教徒的比重)和种族构成代表"文化(culture)"。

建模者还要假设这个模型中自变量和因变量的关系是线性且直接的。模型中的变量关系也可以设定为线性且间接的(如交互【Interaction】项)、非线性且直接的(通过一个对数转换【Logarithmic Transformation】)以及非线性【Nonlinear】且间接的。

如何评估模型?它们能够准确地解释因变量的方差吗?如果我们使用统计学来估计模型,就可以使用统计标准来评估模型。模型变量的性质决定了该用什么标准来评估。*R* 方【*R*-Squared】统计量是通常被用来评估多元回归【Multiple Regression】模型的统计量。对数似然(log-likelihood)是另一种评估多元模型表现的方法——在这里就是最大似然估计【Maximum Likelihood Estimation】。

模型不只是用来回答那些简单的双变

量关系的问题,它可以处理的问题比这广泛得多。模型有助于社会科学家解释、预测和理解社会现象。它们在完成上述三项任务中的有效度也可以得到评估。

——Ross E. Burkhart

(李洋译 高勇校)

参考文献

Asher, H. B. (1983). *Causal modeling* (2nd ed., Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-003). Beverly Hills, CA: Sage.

Blalock, H. M., Jr. (1969). *Theory construction: From verbal to mathematical formulations.* Englewood Cliffs, NJ: Prentice Hall.

Lave, C., & March, J. G. (1978). *An introduction to models in the social sciences.* New York: Harper & Row.

Random House. (1997). *Random House Webster's college dictionary.* New York: Author.

Schrodt, P. A. (2002). Mathematical modeling. In J. B. Manheim, R. C. Rich, & L. Willnat (Eds.), *Empirical political analysis: Research methods in political science* (5th ed.). New York: Longman.

调节(Moderating)

参见调节变量【Moderating Variable】。

调节变量(Moderating Variable)

调节变量表示能改变自变量 X 对因变量 Y 影响的过程或因子。调节变量可以采用"指示变量(indicator variable)"(0/1 取值)、分类变量或者连续性变量的形式。对 X 作用于 Y 的效应进行调节,通常应用在理论化较为成熟的研究领域中,例如,在个体劳动力供给这个领域中,已经建立起了人力资源和劳动力市场分割理论等完善的理论体系。可以将反映这两个理论的变量纳入多元回归方程中,并发展出两个阶段(第一个阶段是劳动力参与,第二阶段是工资决定)的路径分析。模型中的路径涉及对两个回归的回归系数相乘,得到 X 通过 Z 作用于 Y 的调节效应,并将上述效应与 X 对 Z(假设 Z 是加性可分的)的直接效应进行比较(Bryman & Cramer,1997)。

此外,还可通过检验背景因子,得到每个 X 对最终结果 Y 主要影响的调节效应。

调节变量还可对条件效应(contingent effects)进行检验。唯实论者区别了必然效应(necessary effects)和条件效应(Sayer, 2000,p.94)。研究者关注只有在不存在抵消性的(offsetting)、阻碍性的(blocking)、干扰性的因果过程时才会出现的条件效应,会受益匪浅。调节变量可能会成为 X 作用于 Y 的障碍。因而,如果离婚案件事实清楚、没有争议,那么,只要"提交离婚协议"就可以离婚;也就是说,离婚协议只有在满足上述条件时才能引发离婚事件。

目前主要有两种对调节变量的影响进行估计的方法:路径分析【Path Analysis】和交互效应【Interaction Effect】。

路径分析法由两个或两个以上的方程式共同表示一组关系。这些方程可能包含

自变量 X_{ij}、中间变量 Z_{ij} 以及最后的结果变量 Y_{ij}。在每个方程 j 中，变量可能受到其他关联方程中的因素影响。这样的模型就是结构方程模型（structural equation models）。

使用交互效应（Jaccard，Turrisi & Wan，1990）是另一种方法，它在回归方程中引入一个额外项，这个额外项是两个已有变量的乘积。对于分类变量来说，可能会有很多交互项。新的估计值的自由度会减少。

例如，在研究与外出用餐频率相关联的因素时，可通过在 Logistic 回归分析中使用虚拟变量，来测量家庭有年幼孩子和婚姻状况两个因素的交互效应（Olsen，Warde，& Martens，2000）。在外出去餐厅吃饭或者吃民族特色餐的概率上，如果加入（离婚×有不到 5 岁的孩子）这一交互项，就会发现这一交互项大大增加了外出用餐的概率。在英国，与已婚/同居人士相比，离婚人士通常更少外出用餐，这是一个基本关系；但是，如果离婚者家中有年幼孩子，就会抵消这种基本关系，他们会更多地外出用餐。

还有一个简单的方法是使用虚拟变量来控制调节因子的影响；另一个方法是采用多层模型来代表情境（context）性的调节变量的影响（Jones，1997）。

——Wendy K. Olsen

（李洋译　高勇校）

参考文献

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bryman, A., & Cramer, D. (1997). *Quantitative data analysis with SPSS for Windows: A guide for social scientists*. New York: Routledge.

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression* (Quantitative Applications in the Social Sciences, No. 72). London: Sage.

Jones, K. (1997). Multilevel approaches to modelling contextuality: From nuisance to substance in the analysis of voting behaviour. In G. P. Westert & R. N. Verhoeff (Eds.), *Places and people: Multilevel modelling in geographical research* (Nederlandse Geografische Studies 227). Utrecht, The Netherlands: The Royal Dutch Geographical Society and Faculty of Geographical Sciences, Utrecht University.

Olsen, W. K., Warde, A., & Martens, L. (2000). Social differentiation and the market for eating out in the UK. *International Journal of Hospitality Management*, *19*(2), 173-190.

Sayer, A. (2000). *Realism and social science*. London: Sage.

矩（Moment）

在统计分析中，“moment”的意思是矩，通常要与其阶数（order）同时使用。变量的第 k 阶矩，就是变量观测值的 k 次方的平均值。如果 N 表示样本大小，$x_1, x_2, \cdots, x_N$ 表示变量的个案，那么变量的第 k 个矩就是

$$\frac{x_1^k + x_2^k + \cdots + x_N^k}{N}$$

随机变量 X 的第 k 阶矩是 X^k 的期望值或 $E(X^k)$。

在定性研究中，“moment”的意思是时刻，通常用来指某一种社会和历史情境的特定组合，例如，某一社会历史时刻（sociohistorical moment）。

——Tim Futing Liao

（李洋译　高勇校）

定性研究的发展阶段(Moments in Qualitative Research)

20世纪,北美定性研究的历史可分成七个时期。我们将其称为传统期(traditional, 1900—1950),现代期(modernist,1950—1970)或黄金期,类型混杂期(blurred genres,1970—1986),表征危机期(crisis of representation,1986—1990),后现代期(postmodern,1990—1995),一个实验民族志和新民族志的时期、后实验探究期(postexperimental inquiry,1995—2000),以及现在所处的未来期(future)。第七个时期中,"未来期"重点关注道德话语;艺术、文化和科学重新携手并进;神圣文本性(sacred textualities)得到发展。第七个发展阶段要求社会科学和人文学科成为对民主、种族、性别、阶级、国家、全球化、自由和社区进行批判性对话的场所。

传统期

第一个阶段是传统时期,它始于1900年代早期并一直持续发展到第二次世界大战。在这个时期,定性研究人员对田野经历进行"客观的"、开拓式的记述,这反映了实证主义【Positivism】的科学范式。他们着力于在其著作中提供有效、可靠和客观的解释。被研究的对象通常来自异域。

现代期

现代时期或者说第二个发展阶段,是在传统时期的经典著作基础上发展而来的。社会实在主义【Realism】、自然主义【Naturalism】和横截面式的民族志【Ethnography】仍然受到推崇。它从战后几年一直延展到1970年代,当今仍然存在于许多人的研究中。在这个阶段,很多著作试图把定性方法正规化和合法化。现代期的民族志学者和社会学中的参与观察者【Participant Observation】尝试对重要的社会过程进行缜密的定性研究,包括班级和社会中的越轨行为与社会控制(social control)。这是一个创造性高涨的时期。

类型混杂期

在第三个阶段(1970—1986)"类型混杂"中,定性研究者有一系列研究范式、方法和策略可用于研究中。理论包括符号互动论【Symbolic Interactionism】到建构主义【Constructivism】、自然主义探究、实证主义和后实证主义、现象学【Phenomenology】、常人方法论【Ethnomethodology】、批判马克思主义、符号学【Semiotics】、结构主义【Structuralism】、女性主义【Feminism】和各种人类学范式。应用定性研究【Applied Qualitative Research】的地位逐渐提升,定性研究的政治和伦理问题是引起较大关注的两个议题。研究策略和研究成果的形式多种多样,从扎根理论【Grounded Theory】到个案研究【Case Study】,再到历史研究方法、传记体方法、民族志方法、行为研究方法和临床研究方法。收集和分析经验材料的方法也多种多样,包括定性访谈(开放式或者半结构化)、观察法、视觉分析法、个体经验法和文献法。计算机也已经进入研究中,并在以后的十年充分发展(提供了在线方法),与之相伴的还有解读访谈和文化文本的叙述、内容和符号学方法。格尔茨的两本著作,《文化的解释》(*The Interpretation of Cultures*, Geertz, 1973)和《地方性知识:阐释类学论文集》(*Local Knowledge*, Geertz, 1983),是这一发展阶段开始和结束的标志。

表征危机期

在1980年代中期，定性研究产生了深刻的分裂，称第四个发展阶段为“表征危机期”，这一阶段出现了《作为文化批评的人类学》（*Anthropology as Cultural Critique*，Marcus & Fischer，1986）、《经验的人类学》（*The Anthropology of Experience*，Turner & Bruner，1986）、《写文化》（*Writing Culture*，Clifford & Marcus）等著作。他们在寻求新的真理、方法和表征的模型时，清晰地展现了格尔茨在1980年代早期进行的“文体混杂”式解释的后果（Rosaldo，1989）。

后现代期

第五个发展阶段是实验民族志写作的后现代阶段。民族志学者力争对前一阶段的分裂和危机进行深入的分析理解。人们探索了民族志写作的新方法。写作者努力用新的方法来展现文本中的“他者”。从原先被压制无声的群体中产生了新的认识论，为这些问题提供了解决之道。冷漠观察者的概念已经被摒弃。

后实验探究期及未来期

第六个（后实验）和第七个（未来）发展阶段是我们正在经历的。虚拟民族志（fictional ethnographies）、民族志诗学（ethnographic poetry）和多媒体文本今天都已经习以为常。后实验写作者寻求将其作品与自由民主社会的需求联系起来。许多新近作者正在积极探索一种富于道德性和神圣性的定性社会科学（Clough，1998）。

从简短的历史沿革中得出几个结论，不过如同所有历史一样，上述历史都多少有武断的成分。

第一，不是只有北美学者在努力创造后殖民主义（postcolonial）、非本质主义（nonessentialist）、女权主义、对话（dialogic）和表演（performance）文本，即由人文科学的修辞学和叙事转向而激发出来的那些文本。这是一项国际性的工作，它跨越了科学、人文、修辞、文学、事实、虚拟之间的传统界限。正如阿特金森哈默斯利（Atkinson & Hammersley，1994）所言，这一话语确认了“民族志文本的文学源头，申明了基本的辩证性对话”存在于这些美学和人文运动背后（p.255）。

第二，反思性地讲，这一叙述是有其多元的、历史的、国别的背景。很显然，美国定性研究的历史不能推广到世界的其他地方（Atkinson，Coffey & Delamont，2001）。不是所有的研究人员都信奉一个政治化的、文化性的研究议程，这种议程要求解释性文本能够推进有关社会正义和种族平等的议题。

第三，每个先前的历史发展阶段在当前仍然发挥作用，或者是作为学术遗产，或者是研究人员仍然在遵循或反对的一种实践。定性研究的这种多元和断裂的历史，使得任何一个研究人员都可以把一个研究计划与前面描述的某个历史时期的经典文本联系上（canonical text）。在这一领域中有多个评价标准在相互竞争。

第四，当前定性研究领域中的一个重要特征是选择困境（embarrassment of choices）。有如此众多的范式、调查策略、分析方法可以利用，这是前所未有的。

第五，我们处在一个发现和再发现的阶段，新的观察、解释、论述和写作方法被加以争论和讨论。

第六，定性研究行为不再被视为来自中立的、客观实证主义的视角。阶级、种族、性别和族裔不可避免地形塑了探究的过程，使得研究成为一个多元文化的过程。

——Norman K. Denzin

Yvonna S. Lincoln

（李洋译　高勇校）

参考文献

Atkinson, P., Coffey, A., & Delamont, S. (2001). Editorial: A debate about our canon. *Qualitative Research*, *1*, 5-21.

Atkinson, P., & Hammersley, M. (1994). Ethnography and participant observation. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 248-261). Thousand Oaks, CA: Sage.

Clifford, J., & Marcus, G. E. (Eds.). (1986). *Writing culture*. Berkeley: University of California Press.

Clough, P. T. (1998). *The end(s) of ethnography* (2nd ed.). New York: Peter Lang.

Geertz, C. (1973). *The interpretation of cultures*. New York: Basic Books.

Geertz, C. (1983). *Local knowledge*. New York: Basic Books.

Marcus, G., & Fischer, M. (1986). *Anthropology as cultural critique*. Chicago: University of Chicago Press.

Rosaldo, R. (1989). *Culture & truth*. Boston: Beacon.x

Turner, V., & Bruner, E. (Eds.). (1986). *The anthropology of experience*. Urbana: University of Illinois Press.

单调(Monotonic)

单调是将一组值与另一组值关联起来的函数的属性。用公式来表达,设变量 Y 与变量 X 的函数关系为 $Y=f(X)$。对于任意两个值 X_j 和 X_k,如果 $X_j<X_k$,则 $f(X_j) \leqslant f(X_k)$ 成立,那么,这个函数就是单调增长函数。如果 $f(X_j) < f(X_k)$,那么这个函数就是严苛单调增长(monotonic strictly increasing)函数。如果 $f(X_j) \geqslant f(X_k)$,那么这个函数就是单调递减(monotonic decreasing)函数。如果 $f(X_j) > f(X_k)$,那么这个函数就是严苛单调递减(monotonic strictly decreasing)函数。

例如,考察 5 个个体在变量 X 以及另外 5 个额外的变量 A,B,C,D,E 中的观测值,数据见表 1。

表 1

个体	X	A	B	C	D	E
1	12	44	44	47	47	47
2	13	45	44	46	46	46
3	14	46	45	45	45	45
4	15	47	46	44	44	46
5	16	48	47	43	44	47

由于对任意两个 X_j 小于 X_k 的个体,都有 $A_j<A_k$,因此,A 是 X 的严苛单调增长函数。B 是 X 的单调增长函数,因为对于任意两个 X_j 小于 X_k 的个体,都有 $B_j \leqslant B_k$。C 是 X 的严苛单调递减函数,因为对于任意两个 X_j 小于 X_k 的个体,都有 $C_j > C_k$。D

是 X 的单调增长函数，因为对于任意两个 X_j小于 X_k的个体，都有 $B_j \geq B_k$。最后，E 是 X 的非单调函数，因为它不满足上述任何条件。

——James J. Jaccard

（李洋译　高勇校）

蒙特卡洛模拟（Monte Carlo Simulation）

蒙特卡洛模拟是采用计算机算法【Algorithm】和伪随机数来对统计量进行数学实验的方法。通过这些实验可以用来理解特定条件下统计量的性质，并最终对总体特征进行推论。如果某一个统计量在特定情形下无法基于分析把握其分布特征时，蒙特卡洛模拟是最有效的分析工具。

解释

推论统计【Inferential Statistical】的目的，就是用从总体中抽样的数据计算出统计量 $\hat{\theta}$ 来对总体参数【Population Parameter】θ 进行概率【Probability】陈述。这里的问题是，从同一总体中抽样两个随机样本【Random Sample】，计算出来的统计量几乎不可能相等。例如，假设你想了解你所在大学中20 000名学生的 IQ 平均值 θ，但是你没有时间和资源来对全体学生进行测验。因此，你从中抽取了一个 200 名学生的随机样本，然后计算样本的 IQ 平均得分 $\hat{\theta}$，比如说结果是 123。那么全体学生的 IQ 平均值是多少呢？这并不显而易见，因为，如果你从同样的 20 000 名学生中再次抽取 200 名，他们的 IQ 平均得分这次可能是 117。第三个样本可能是 132，你从全体学生中抽取很多次样本，每次的结果都可能不同。那么，你要怎样做才能从关于样本统计量 $\hat{\theta}$ 的可靠信息中形成关于总体参数（θ）的概率描述呢？

从样本推论到总体的核心是抽样分布【Sampling Distribution】。一个统计量的抽样分布是它在一个从总体中抽取出的随机样本中的取值范围以及与这些值相应的概率。因此，从 20 000 名学生中抽取 200 名学生的 IQ 平均值可能低到 89，也可能高达 203，但是更大的可能性是 IQ 平均值为 120~130。如果我们掌握了样本均值的抽样分布的结构，就可以通过一个既定的样本形成关于总体均值的概率陈述。

在社会科学家从研究生院学到的标准参数统计推论中（如应用广泛的 t-检验和 p 值），我们可以通过数学分析的方法得到统计量的抽样分布信息。例如，中心极限定理【Central Limit Theorem】可以证明 200 名学生随机样本的抽样分布服从正态分布、期望值为总体均值、标准差【Standard Deviation】为总体 IQ 值的标准差除以 200 的平方根。但是有时关于某一统计量的分析性的分布理论不存在，或者使用参数统计理论所需要的假设不成立。在这些情况下，就可以用蒙特卡洛模拟来经验性地估计统计量的抽样分布。

蒙特卡洛模拟通过从一个由模拟数据构成的已知的“伪总体”中抽取大量随机样本来对统计量的抽样分布进行经验估计，以把握统计量的性质。其基本理念很简单：某一统计量的抽样分布是它在给定总体中所可能取值的密度函数，那么，从总体中抽取很多样本，然后观察这些样本中这一统计量的相对频率分布就可估计出此抽样分布。对社会科学家来说，不太可能多次抽取出真

实的数据，因此，我们使用在相关方面与真实情况相似的人工生成数据来替代真实的样本。

蒙特卡洛模拟的基本步骤如下：

（1）象征性地定义一个用来生成随机样本的伪总体。通常需用到一种计算机算法，它能够以精心设定的方式生成模拟数据。

（2）从伪总体中进行抽样（得到一个伪样本），抽样方式与要研究的统计条件相同，例如，采用同样的抽样方案、样本量等。

（3）计算并保存从伪样本中得到的 $\hat{\theta}$。

（4）将步骤 2 和步骤 3 重复 t 次，其中 t 是试验次数。通常来说，t 要很大才行，一般要达到 10 000 或更大。这就可以通过计算机程序中的循环算法来实现。

（5）构建一个 $\hat{\theta}_t$ 值的相对频次分布。这就是在伪总体和抽样程序设定的条件下，$\hat{\theta}$ 的抽样分布的蒙特卡洛估计值（Monte Carlo estimate）。

（6）评估或使用估计出的抽样分布来理解总体取值为 θ 时 $\hat{\theta}$ 的性质。

因而，蒙特卡洛模拟就是通过编写和运行能够人工生成结构性的随机数据的计算机程序，来对某一统计量的特征进行数学实验。这一过程中两个关键步骤是：（a）在计算机算法中设定伪总体；（b）应用估计出来的抽样分布来更好地理解社会过程。

用计算机算法来设定伪总体，必须对要模拟的社会过程有清晰的理解。此类计算机算法必须和要模拟的社会过程一样复杂。大多数社会过程的理论模型都包含确定性部分（deterministic component）和概率部分或随机部分（probabilistic or random component）。为确定性部分写计算机代码通常是相对容易的。模拟随机分量通常需要多加思量。这是一种恰当的做法，因为社会过程的随机部分很大程度上决定了统计量抽样分布的构造，是统计推论中的头等要事。正因为这项技术中随机部分很重要，它才会用欧洲著名赌场的名字来命名。

这些随机部分是通过模拟过程中各个变量的分布来进入对社会过程的蒙特卡洛模拟的。你在写模拟程序时，对这些分布的设定要基于要建模的这一社会现象的理论和经验证据，以及你要进行的实验特征。在理论上，一个社会变量可以呈现无限多的分布形态，但是统计学家只对其中相对较少的分布能够识别并发展出了相关理论。大多数关于蒙特卡洛模拟的著作，包括下面将要提到的这些著作，都详尽地讨论了如何生成具有不同分布的变量，以及用这些去模拟什么样的可观察变量才最好。

应用

蒙特卡洛模拟基于数值实验的传统，它至少可以追溯到圣经时代，但是，直到 20 世纪中叶，人们才在量子物理学和电话网等完全不同的领域中重新激发出对蒙特卡洛模拟的兴趣。但是直到最近这项技术才开始广泛应用，原因在于 20 世纪末计算机首次具有的巨大能力。因此，社会科学家对蒙特卡洛模拟的一般性应用才刚开始（但是相关应用增长迅速，1987 年在社会科学引文索引 SSCI 中搜索蒙特卡洛仅有 14 篇文献，2001 年这一数字上升到 270 篇）。当前蒙特卡洛模拟最重要的应用主要包括：

（1）对数学分析难以处理的新统计量（或者统计量的合并）的抽样分布进行估计。

（2）评估某种经验分析在各种可能的总体条件下（如违背参数推论的前提假设

时)的敏感度。

(3)用于马尔科夫链蒙特卡洛【Markov Chain Monte Carlo】技术中,这种技术最近将贝叶斯统计推论引入主流社会科学中。

关于社会科学中蒙特卡洛模拟的应用,可参考Evans,Hastings,& Peacock,(2000);Mooney(1997);Robert & Casella,(2000);以及Rubinstein,(1981)。

——Christopher Z. Mooney

(李洋译 高勇校)

参考文献

Evans, M., Hastings, N., & Peacock, B. (2000). *Statistical distributions* (3rd ed.). New York: John Wiley.

Mooney, C. Z. (1997). *Monte Carlo simulation.* Thousand Oaks, CA: Sage.

Robert, C. P., & Casella, G. (2000). *Monte Carlo statistical methods.* New York: Springer-Verlag.

Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo method.* New York: John Wiley.

马赛克图(Mosaic Display)

马赛克图是对交叉分类列联表【Contingency Table】中不同单元格的频次进行绘图,它将不同单元格的频次 $n_{ijkl...}$ 绘制成马赛克图中用以填充的"砖块"面积。马赛克拼图是基于以下理念产生的:这对一个二维表格来说,单元格概率 $P_{ij} = n_{ij}/n_{++}$ 总能表达为 $P_{ij} = P_{i+} \times P_{j|i}$,即用长方形的宽和高来表示。

因此,如果一个单位面积首先根据变量边际概率的比例 p_{i+} 进行划分,然后再根据条件概率 $p_{j|i}$ 进行细分,那么,其中砖块的面积就为 p_{ij},而且在独立条件满足时($p_{ij} = p_{i+} \times p_{+j}$时)砖块就会对齐。这种可视化方式的历史既悠久又有趣(Friendly,2002)。

最近马赛克拼图又被重新引入统计图中(Hartigan & Kleiner,1981),而且马赛克拼图的基本理念得以一般化到 n 维表格中,方法是不断采用贝叶斯准则(即 $p_{ijkl...} = p_i \times p_{j|i} \times p_{k|ij} \times \cdots \times p_{n|ijk...}$)对面积进行递推性的划分。

这种可视化列联表的方法之所以得到强化(Friendly,1992,1994,1998),其目的是适应任意的对数线性模型,通过颜色和阴影显示模型的残差(对总体皮尔逊或似然比卡方的贡献),以及对变量类别进行重新排序,使阴影图案显示变量之间的关系属性。这样的话,马赛克拼图已经成为对 n 维列联表形式中类别数据进行可视化和分析的主要图形工具(Friendly, 1999, 2000; Valero, Young,& Friendly,2003)。

图1中的二向表和三向表显示了头发颜色、眼睛颜色和性别之间的关系。在这些表格中,每个砖块面积与观测单元格的频次成比例;此外,根据特定对数线性模型的残差,砖块被涂上了不同的阴影,以此显示剩余的关系模式(用颜色更为有效,分别用蓝色和红色阴影表示正向和负向残差)。在左手边的二维图中,其表格拟合一个独立模型;相对角落位置的阴影图案表明深色(浅色)头发与深色(浅色)眼睛相关联。

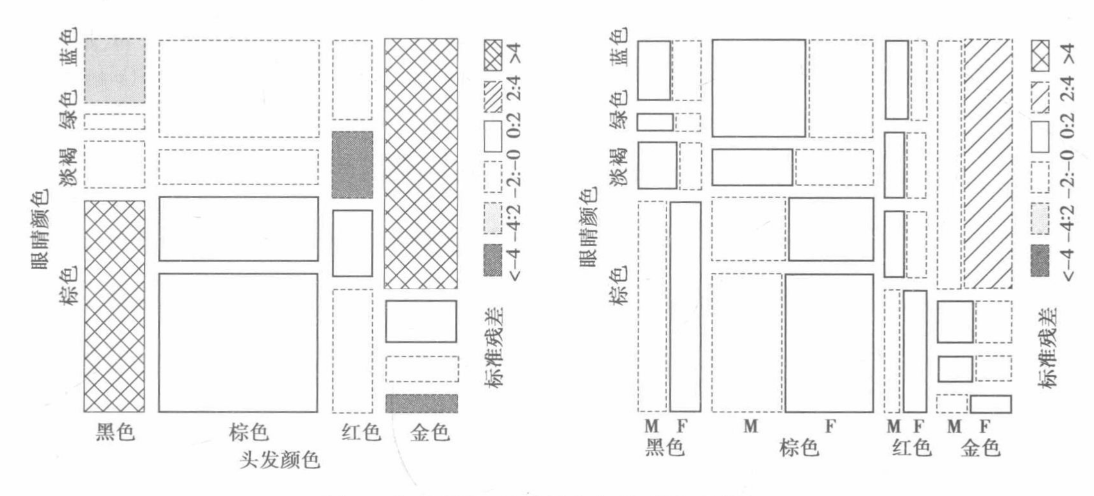


图1 头发颜色、眼睛颜色和性别的马赛克图

注意:左为头发颜色和眼睛颜色的二维表格;右为三维表格,按性别划分。

在右边的面板中,这个三维表格拟合了一个关于[头发颜色][性别]的对数线性模型,可以看出,头发颜色和眼睛颜色的变量组合独立于性别。只有一对单元格的残差大到可以用阴影表示——蓝眼金发女人——在这个独立模型下,就会有女性较多而男性很少的情况。

最近,马赛克图在多个方面取得了进展:(a)如同连续变量中的散点图矩阵和协同图(coplots)一样,离散变量可以绘制出马赛克图矩阵和局部马赛克图(Friendly,1999,2000);(b)可以用几种高度互动、动态性的操作方式来进行探索性研究和"视觉性拟合(visual fitting)";(c)可以将相同的视觉理念应用到嵌套分类或树结构(tree structures)(Shneiderman,1992)中去。

——Michael Friendly

(李洋译 高勇校)

参考文献

Friendly, M. (1992). Mosaic displays for loglinear models. In American Statistical Association (Ed.), *Proceedings of the Statistical Graphics Section* (pp. 61-68). Alexandria, VA: American Statistical Association.

Friendly, M. (1994). Mosaic displays for multi-way contingency tables. *Journal of the American Statistical Association*, *89*, 190-200.

Friendly, M. (1998). Mosaic displays. In S. Kotz, C. Reed, & D. L. Banks (Eds.), *Encyclopedia of statistical sciences* (Vol. 2, pp. 411-416). New York: John Wiley.

Friendly, M. (1999). Extending mosaic displays: Marginal, conditional, and partial views of categorical data. *Journal of Computational and Graphical Statistics*, *8*(3), 373-395.

Friendly, M. (2000). *Visualizing categorical data.* Cary, NC: SAS Institute.

Friendly, M. (2002). A brief history of the mosaic display. *Journal of Computational and Graphical Statistics*, *11*(1), 89-107.

Hartigan, J. A., & Kleiner, B. (1981). Mosaics for contingency tables. In W. F. Eddy (Ed.), *Computer science and statistics: Proceedings of the 13th Symposium on the Interface* (pp. 268-273). New York: Springer-Verlag.

Hofmann, H. (2000). Exploring categorical data: Interactive mosaic plots. *Metrika*, *51*(1), 11-26.

Shneiderman, B. (1992). Tree visualization with treemaps: A 2-D space-filling approach. *ACM Transactions on Graphics*, *11*(1), 92-99.

Theus, M. (1997). Visualization of categorical data.

In W. Bandilla & F. Faulbaum (Eds.), *SoftStat'97: Advances in statistical software* (Vol. 6, pp. 47-55). Stuttgart, Germany: Lucius & Lucius.

Valero, P., Young, F., & Friendly, M. (2003). Visual categorical analysis in ViSta. *Computational Statistics and Data Analysis*, *43*(4), 495-508.

变化者-不变者模型(Mover-Stayer Models)

变化者-不变者(Mover-Stayer,MS)模型是马尔科夫链模型【Markov Chain Model】的一种扩展,用来处理总体中一种特殊类型的不可观测的异质性。总体被认为包括两个无法观察的组别(潜在类别):变化概率为零的不变者和遵循常规马尔科夫链过程的变化者。MS 模型的早期文献包括布吕芒、科根和麦卡锡(Blumen, Kogan, & McCarthy, 1955)和戈德曼(Goodman, 1961)。在不同的应用领域中,这个模型有不同的名称。在生物医学领域,人们可能会遇到长期幸存者这个词,意思是在研究周期内死亡概率或者经历其他某种事件的概率为 0 的群体。在营销研究中,人们使用品牌忠诚群体与品牌转换群体这两个相对的概念。它与用于二项或泊松计数变量的零膨胀模型(zero-inflated)也有很强的关联。

假设对类别变量 Y_t 在 T 时点进行测量,其中 t 表示一个特殊的时点,$1 \leq t \leq T$;D 表示 Y_t 的不同水平数;y_t 是 Y_t 的一个特定的水平,$1 \leq y_t \leq D$。比如说,这可以是每隔 6 个月测量的被访者的政党偏好。向量标记 **Y** 和 **y** 用来表示一个完整的回答模式(response pattern)。

MS 模型可以视为混合马尔科夫模型的一个特例,这样理解起来比较容易。在标准的马尔科夫链模型中,添加一个离散的无法观察的异质性成分,就得到了混合马尔科夫模型。令 X 表示含 C 个水平的离散潜变量。一个潜在类别可以用 x 表示,$x = 1, 2, \cdots, C$。混合马尔科夫链模型可以用公式表示为

$$P(\mathbf{Y}=\mathbf{y}) = \sum_{x=1}^{C} P(X=x) P(Y_1=y_1 \mid X=x) \times \prod_{t=2}^{T} P(Y_t=y_t \mid Y_{t-1}=y_{t-1}, X=x)$$

可以看出,潜在类别之间在原始状态和转换概率上存在差异。如果 $C=2$ 且对于每个 y_t 来说都有 $P(Y_t=y_t \mid Y_{t-1}=y_t, X=2)=1$,得到的就是 MS 模型。换句话说,在潜在类别 2 即不变者类别中,停留在同一状态的概率等于 1。要注意,$P(X=x)$ 是处于变化者与不变者类别的人数比例,$P(Y_1=y_1 \mid X=x)$ 是这两组处于各种状态的初始分布。

为了识别 MS 模型,我们需要至少 3 个时点,但是并不需要假设迁移链中的过程是稳定的。在有三个时点($T=3$)和两个状态($D=2$)的情形下,非稳态(nonstationary)的 MS 模型是一个饱和模型。

MS 模型加入限制条件后,就可以得到独立变化者-不变者模型(independence-stayer model),这里变化者的行为是随机的;也就是说,$P(Y_t=y_t \mid Y_{t-1}=y_{t-1}, X=1) = P(Y_t=y_t \mid X=1)$。MS 模型的另一个类型涉及吸收态(absorbing states)的概念,如初婚和死亡。在这种情况下,每个人的起始状态都是状态 1,$P(Y_1=1 \mid X=x)=1$,但是变化者一旦变化以后就无法再变回去了,即 $P(Y_t=2 \mid Y_{t-1}=2, X=1)=1$。

从混合马尔科夫链模型的公式中还可推出几种扩展形式,例如,可以包括不同种

类的变化者,或者可以用来修正测量误差的潜马尔科夫链模型【Latent Markov Model】。另一种可能的扩展则是引入协变量,这样就得到了一个与用于计数变量的零膨胀模型相类似的模型。

目前有两种程序可以用来估计 MS 模型及其各种扩展模型:PANMARK 和 LEM。

——Jeroen K. Vermunt

(李洋译　高勇校)

参考文献

Blumen, I. M., Kogan, M., & McCarthy, P. J. (1955). *The industrial mobility of labor as a probability process.* Ithaca, NY: Cornell University Press.

Goodman, L. A. (1961). Statistical methods for the mover-stayer model. *Journal of the American Statistical Association*, *56*, 841-868.

Langeheine, R., & Van de Pol, F. (1994). Discrete-time mixed Markov latent class models. In A. Dale & R. B. Davies (Eds.), *Analyzing social and political change: A casebook of methods* (pp. 171-197). London: Sage.

Vermunt, J. K. (1997). *Log-linear models for event histories.* Thousand Oaks, CA: Sage.

移动平均数(Moving Average)

移动平均过程是一个随机【Random】震荡(shock)或扰动在某个时间序列【Time-Series】中的产物。如果因变量中的扰动模式可以表示为当前扰动和时滞随机扰动的加权平均数时,这就是移动平均过程。随机震荡有当下效应(immediate effects),同时其效应会经过一段时间之后才消失,因此,它兼有衰减效应(discounted effects)。在某种意义上移动平均数是白噪声【White-Noise】误差项的线性组合。随机震荡对因变量的影响需要多少个周期才能消失,这被称为移动平均过程的阶【Order】,用 q 表示。随机震荡效应的长度并不取决于震荡发生的时间,也就是说,移动平均过程是独立于时间的。

均值 u 为 0 的 q 阶移动平均过程的数学表达式是:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

以一阶移动平均过程(即 $q=1$)为例:t 期的随机震荡会在 t 期和 $t-1$ 期产生影响,随后不再有影响。相隔一期的观测值之间是相关的,但相隔多期的观测值之间并不相关。这一过程仅有一期的记忆。这一过程的阶数被称为这一过程的记忆。有一个例子,是关于艾奥瓦州猪饲料价格的新闻报道的效应。这一随机震荡的效应在第一天就能感受到,然后在 1 天或更多天(q 阶)不断衰减。几天后,其效应将会消失,导致这一序列回归于其均值。

对移动平均过程的识别需要考察自相关【Autocorrelation, ACF】函数和偏自相关(partial autocorrelation, PACF)函数。对于一个一阶移动平均过程来说,自相关函数会有达到一个峰值,然后突然削弱,函数的剩余时滞会呈现白噪声的特性;一阶移动平均过程的 PACF 值会一直消退直至无穷。图 1 显示了一阶移动平均过程的自相关函数。q 阶移动平均过程的 ACF 会从时滞 1 到时滞 q 呈现出显著的峰值,但在时滞 q 之后突然削弱;PACF 则会一直消退直至无穷。统计包如 SAS 和 SPSS 能生成自相关图(correlograms)来快速评估其自相关函数和偏自相关函数。在社会科学中,高阶移动平

均过程是很罕见的。

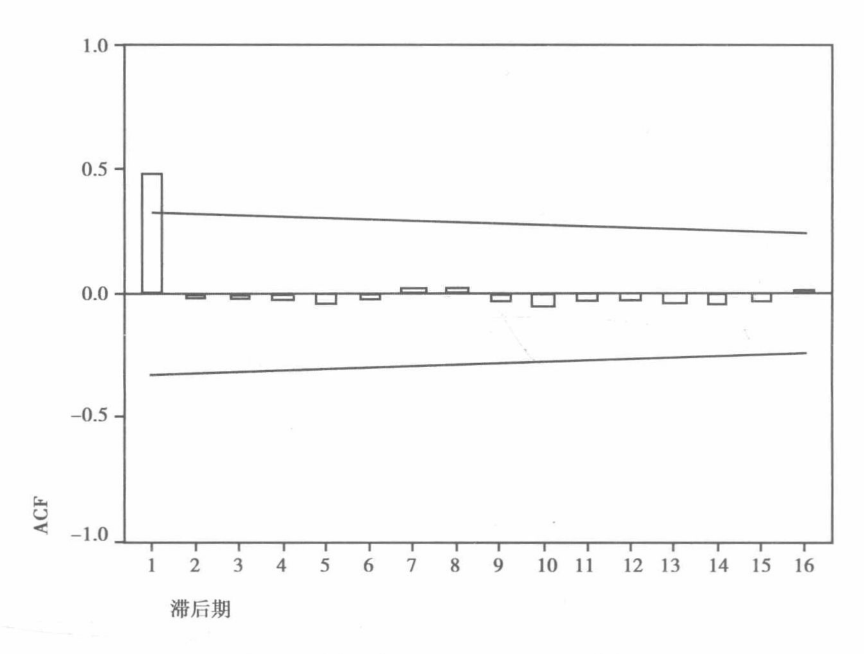


图 1 理论自相关(ACF):MA(1)过程

移动平均过程可以与序列相关(serial correlation)和趋势【Trend】同时出现。自回归差分移动平均模型(Autoregressive Integrated Moving Average,ARIMA)可以用于对兼具上述过程的时间序列建模,因为它综合了自回归过程、积分或差分过程(用来去除趋势)、移动平均过程。此外,移动平均过程也可表达为无穷阶的自回归过程。反之,无穷阶自回归过程也可表达为一阶移动平均过程。这一点很重要,因为它可以把复杂的过程简约地表示为其他过程的低阶形式。

——Andrew J. Civettini

(李洋译 高勇校)

参考文献

Cromwell, J. B., Labys, W., & Terraza, M. (1994). *Univariate tests for time series models.* Thousand Oaks, CA: Sage.

Ostrom, C. W., Jr. (1990). *Time series analysis.* Thousand Oaks, CA: Sage.

Mplus

Mplus 是一种统计软件,它可以用来估计结构方程模型【Structural Equation Modeling】,也可以进行整合随机效应因素【Random-Effects Factor】和潜在类别分析【Latent Class Analysis】的分析。分析数据可以是横截面的,也可以是纵贯性的;可以是单层次的,也可以是多层次的。更多关于软件的信息请参见相关网址。

——Tim Futing Liao

(李洋译 高勇校)

多重共线性(Multicollinearity)

两个自变量【Independent Variable】之间的共线性或者线性回归【Linear Regression】中多个自变量之间的多重共线性,意味着这些变量之间存在线性关系。如果用向量来表示变量,完全共线性意味着两个向量是在同一条线上,或者更一般地讲,如果有 k 个变量,则这些向量位于低于 k 维的子空间中(即一个向量是其他向量的线性组合)。变量 **X** 之间存在完全线性关系(exact linear relationship)的结果就是叉差(cross-product)矩阵 **X′X** 是奇异矩阵(即行列式等于0),所以它没有逆矩阵。实践中很少出现完全线性关系,但是社会现象之间的互依性(interdependence)可能导致"近似"线性关系的出现。这种现象被称为多重共线性。

对诸如主成分【Principal Component】和因子分析【Factor Analysis】这样对互依性(interdependence)的分析中,出现多重共线性是好事,因为它们的目标是简化维度;而在诸如多重回归分析【Multiple Regression Analysis】这样的对依存结构(dependence structures)的分析中,自变量之间的多重共线性较麻烦。这里我们将多重共线性放在回归分析的背景中进行分析。

结果

假设对失业工人的幸福感原因进行调查(幸福感按尺度从0~100进行测量)。许多解释变量【Explanatory Variable】被纳入考察范围,例如,年收入(年收入越低幸福感越低)、失业时间(失业周数越多幸福感越低)以及其他变量。在这些自变量中,有些变量之间的相关程度如此之高,以至于多元回归分析得出的结论会受到影响。在一个多因模型(multicausal model)中,高度的多重共线性是不受欢迎的,因为标准误会变大,置信区间【Confidence Intervals】会变宽,进而会影响显著性检验【Significance Testing】和统计推论【Statistical Inference】,虽然最小二乘【Ordinary Least Squares】估计依然是无偏的。与方差【Variance】和协方差【Covariance】一样,多重共线性也是样本的性质。这意味着分析可能是不可靠的,并且对离群值【Outlier】和极端案例较为敏感。

在社会科学研究中,因果模型中自变量存在相关关系是老问题了。在多元回归分析中,这被称为多重共线性(以向量来表示的预测变量几乎重合于一条直线),而在方差和协方差分析【Analysis of Variance and Covariance】中,人们倾向于使用非正交性这个词语(因子响应模型的因子不互相垂直),多重共线性和非正交性这两个词实际上是同义词。

让我们重点看一下存在多重共线性问题的分析中会发生什么奇怪的事情。肯德尔和斯图尔特(Kendall & Stuart,1948)提供了一个极端的例子,一个因变量 Y 和两个自变量 X_1 和 X_2,其中,Y 与 X_1、Y 与 X_2 的相关性很弱(其中一个甚至为0),X_1 和 X_2 的相关性很强(0.90,高度多重共线性),而多重相关系数 $R_{y \cdot 12}$ 近似于1。从这个代表性案例中可以看出多重共线性的典型后果:每个参数单独检验结果不显著,但是 R 方【R-Squared】值很高。标准化系数(样本方差的直接函数)通常不会大于1,因为其平方值表示了能够解释的方差比例,但是在这里标准化系数超过了2;结果两个变量似乎每个都能解释400%的方差,但实际只有100%需要解释!

检查

我们如何才能知道分析是否受到了强

多重共线性的影响呢?

多数统计软件包(如 SPSS, SAS 和 BMDP)都有"共线性诊断"部分。标准输出结果中会包括与其他自变量的相关系数、从 X_j 对所有其他自变量的辅助回归中得到的多重相关系数、容忍度【Tolerance】($1-R_j^2$)、方差膨胀因子(variance-inflation factors)[$1/(1-R_j^2)$]和特征值分析(参见 Belsley, Kuh, & Welsch, 1980),从中可以看出哪些自变量是有问题的。这些诊断方法比某些教科书所建议的主观划定相关系数的一个临界线(如 0.60)要更合理。此外,这些诊断法也可与回归系数的显著性检验【Significance Testing】(或者置信区间的计算)结合起来使用。此外,不能拿 X 变量之间的相关性作为判断多重共线性是否存在的充分证据,不要忘记必须在多变量的背景中评估相关性问题;也就是说,它可能受到虚假相关、抑制变量(suppressor variables)或其他类似因素的影响。

解决方法

第一种方法是忽略多重共线性问题。只有在研究者对于涉及多重共线性的各个变量的效应都不感兴趣时,这种办法才是合理的。

第二种显而易见的方法,就是删除其中一个高度相关的变量。但是研究者通常发现,从理论上讲这种办法是说不通的。拿前面提到的例子来说,即使年收入和失业时间在统计上相关,但它们都对幸福感给出了各自完全不同的解释。

第三种可能的方法是由计量经济学家提出的,就是扩大样本容量。通过获取更多的数据,当前样本及其性质得以改善。但是这种方法在大多数情况下并不现实,因为这需要额外的时间和经费。

第四种方法是从一个变量中去除它与一个共线变量的公共方差。在当前的例子中,这需要研究人员保留年收入这一变量,但仅保留失业时间对年收入进行回归后的残差。这种方法可以用于所有涉及的变量上。

第五种方法是生成一个共线变量的综合指数(summary index),前提是研究人员对观察变量的单独效应不感兴趣。但是这种方法忽略了测量误差和其他可能的因素。

更为成熟的一种方法是使用主成分分析法(Principal Components Analysis, PCA)。对所有的自变量(数目通常很大)运行主成分分析法,可能还要进行方差最大旋转【Varimax Rotation】,这样得到的成分得分被用到进一步的回归分析中,例如,用幸福感作为因变量,用不同的成分作为自变量。这样自变量之间就彼此垂直,避免了多重共线性。对新成分的解释可能是一种挑战,但是,这个问题可通过只对共线变量进行 PCA 分析来避免。

最后,当研究者既打算在回归模型中保留高度相关的变量 X_1 和 X_2,又要进行有充分科学根据的分析时,我们推荐采用岭回归(如 Hoerl & Kennard, 1970)和贝叶斯回归(如 Lindley & Smith, 1972),但它们也无法提供 100%令人满意的解决方案。

岭回归是通过在 OLS 估计中引入一些偏差来对多重共线性问题进行处理的一种尝试。岭回归估计的标准误较低,但是它们可能略微有偏差。在 $\mathbf{Y}=\mathbf{X\beta}+\varepsilon$ 回归中,系数估计为 $\hat{\beta}^*=(\mathbf{X'X}+k\mathbf{I})^{-1}\mathbf{X'y}$,其中主对角线($\mathbf{X'X}$)加上了一个很小的增加值 k。这样,整个方程更具有正交的特征。但是较小值 k 的选择仍然是主观的。k 值越大,标准误就越小,但是估计偏差也就增加得越多。岭回归在当前主流的统计软件中都可以找到。

与岭回归类似,贝叶斯回归(Bayes regression)也尝试对多重共线性问题进行修正。岭回归中,较小的增加值 k 是要依靠研究人员来主观选择的,而在贝叶斯回归中它们是根据数据估计出来的。贝叶斯方法的

缺点是，它要预设出先验分布（priori distribution）。

——Jacques Tacq

（李洋译　高勇校）

参考文献

Belsley, D. A., Kuh, E., & Welsch, R. (1980). *Regression diagnostics.* New York: John Wiley.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression. *Technometrics*, *12*(1), 55-67, 69-82.

Kendall, M. G., & Stuart, A. (1948). *The advanced theory of statistics.* London: Griffin.

Lindley, D. V., & Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society*, *34*, 1-41.

多维度量（Multidimensional Scaling, MDS）

多维度量是指通过图示用空间中一组点之间的关系来表示一组数据结构的一些模型。多维度量可以适用于许多种类的数据，可以使用多种不同的模型，也可以对于测量层次有多种不同的假设。

最简单的例子是，一个数据矩阵给出了一组对象之间的相似性（或相异性【Dissimilarity】）信息，那就可以用低维空间中对应点的接近程度（或者距离）来表示上述信息。假设有一组数据，可以解释为“距离（distances）”，那么就可以发现生成距离的图点。

例如，在一项关于人们如何对药物进行分类的研究中，药物的清单可以从药物使用者或非使用者的样本中间接获得；选择28种药物以进行自由分类（free-sorting）试验，并用同时出现的频次来进行相似性测量。如图1所示，数据按顺序排列在两个维度上。

多维度量的最终构形：维度1和维度2　Stress1=0.097

1.酒精
2.安非他命
3.阿司匹林
4.巴比妥
5.咖啡因
6.大麻
7.巧克力
8.可卡因
9.止咳药水
10.纯度可卡因
11.迷幻药
12.石墨
13.胶水
14.海洛因
15.止泻剂
16.胰岛素
17.克他命
18.二乙基麦角酰胺
19.致幻蘑菇
20.美沙酮
21.苯环己哌啶
22.盘尼西林
23.芳香剂
24.百忧解
25.类固醇
26.羟基安定
27.烟草
28.威而刚

图1

在数据理想的情况下，相异性数据（dissimilarity data）和解的距离（distances of the solution）之间是完全对应的，但是如果数据不理想，拟合度就要用标准化 STRESS（残差平方和）值来表示，它测量的是拟合劣度（badness of fit）。在该例子中，拟合较好（$stress_1 = 0.097$），比随机期望值小了 3 倍。从解的构成来看，一共生成 3 个高度分化的群组（通过独立多层聚类确认），分别是轻微兴奋剂类药物（核心药物：烟草、咖啡因）、强烈消遣类（核心药物：可卡因、海洛因）、家庭处方药类（核心药物：阿司匹林、盘尼西林）。

多维度量可以是探索性的（仅给出数据的一个有用且易于理解的可视化方式）或者解释性的（用几何方式表达出数据的结构，其中模型假设被用来表示数据的生成方式）。与其他多变量方法相比，MDS 模型通常是自由分布（distribution free）的，对数据结构要求较为保守，不容易受到非系统缺失数据的影响，并且可以广泛用于各种测量，模型解通常也很容易进行解释。主要弱点是它更容易忽略 STRESS 统计量的抽样特征，可能得到的是局部极小解（local minima solutions），无法表现出因果模型（causal models）的非对称性。

基本多维度量模型

MDS 的基本类型分析的是二维的单模数据（如相关矩阵或其他非相似/相似测量），使用的是欧氏距离模型（Euclidean distance model）。在最早的"经典（classical）"的计量（metric）模型中，MDS 将"距离（distances）"转换为标量乘积，然后将其因子化或者分解成一组低维空间中的位置点［因此也称"最小空间分析（smallest space analysis）"］。1962 年，罗格·谢泼德提出了第一个非计量（nonmetric）模型，他表明即便数据仅仅只是定序的，如果数量足够，仍然可以进行计量转换（metric recovery），他同时给出了首个可迭代的计算机程序来实现他的想法。克鲁斯卡尔（Kruskal, 1964）在此基础上增加了最小二乘（ordinary least squares, OLS）的统计内容；塔卡宁、扬和德莱乌（Takane, Young, & Deleeuw, 1977）发展出经常使用的交替最小二乘程序（alternating least squares program, ALSCAL）；同时，采用概率最大似然法【Maximum Likelihood】的 MDS 也相继得以发展（Ramsay, 1977）。

多维度量的各种类型

根据以下方面，可以区分出多维度量的各种不同类型（Coxon, 1982）：

（1）数据，其"形状"可以通过数据的维度（行、列、"第三维"重复观察）和模（不同对象组的数目，如变量的数目或被试组的数目）来进行描述。

（2）转换（重构标度［rescaling］）函数，或者测量层次【Level of Measurement】，它明确设定了所求解的数据属性在多大程度上能得到充分表示。最重要的是要区分非计量的（定序或更低层次的）和计量的（定距的和定比的）尺度层次。

（3）模型，通常是欧氏距离模型，但是也可以采用较简单的 Minkowski 度量（Minkowski metrics），例如，城区距离（City-Block）以及不同的复合函数（composition functions），如向量、标量乘积或者因子【Factor】模型。

二维、单模数据

任意非负、对称性的判断或测量都可以成为基本模型的输入量：频次、计数、相似性的等级/次序（ratings/rankings of similarity）、同时出现（co-occurrence）、搭配（collocation）、

混淆率(confusion rates)、关联量度等。有 12 个或者更多个点,才能得到稳定的二维解(two-dimensional solution)。非计量分析使用距离模型(distance models);计量模型也可以使用向量/因子模型。通过随机配置的模拟研究所得到的 STRESS 值,可以成为评估 STRESS 值的标准。

二维、双模数据

矩形数据矩阵(通常被试为行,变量为列)包括剖面数据(profile data)或者偏好等级/次序(preference ratings/rankings)数据。用距离(未折叠)或者向量模型来生成一个以行列各元素为点的联合双标图【Biplots】。向量模型(MDPREF)形式上与简单的对应分析【Correspondence Analysis】完全一致,用单位向量(unit vectors)表示行元素。由于这些数据是行条件的(row conditional),因此,对所求解的组内关系进行解释时要特别留意。如果例子中的分类数据直接以(个体×对象矩阵,单元格值表示类别数目)的形式录入,那么类似 MDSORT 这种程序可能会同时代表对象和类别。

三维数据

"第三维"可以是数据中的不同个体、不同时空点、不同分组、不同试验条件和不同方法等。最常见的形式是三维双模数据(二维单模矩阵的叠加),在计量加权距离模型(metric weighted distance model)中,这被称为个体差异度量(individual differences scaling, INDSCAL)(Carroll & Chang, 1970)。在个体差异度量的解中,研究对象通过组空间(其维度固定且很容易进行解释)表现出来,第三维("个体")的每个元素有一组非负的维度权重,根据这些权重个体有差异的收缩或扩张组空间(group space)的各个维度,以形成"私用空间(private space)"。因此,个体差异可用公共空间(commonspace)的一组显著权重(salience weights)来表示。权重散布在不同的"被试空间(subject space)"中,各点之间的角间距表示其截面分布的紧密程度,而与起点的距离大体上表示已经解释了的方差。回到例子中,如果个体被分成不同的子组(如性别×作用),并在每个组内分别计算出共生矩阵(co-occurrence matrix),那么这些矩阵就构成了三维、双模的数据,并且通过 INDSCAL 进行的标量计算将会得到下列信息,即根据以药物构成为基础的各个维度的重要性(权重)来判断,不同组之间的差别有多大。

其他形式的 MDS 可以用于其他类型的数据(如表格和三元组)、转换[参数图(parametric mapping)、幂函数(power functions)]和其他模型[非欧氏度量、简单结合模型(simple conjoint models)和离散模型,如加性树(additive trees)],并广泛应用到三维模型和混合模型(hybrid models)中,比如,CLASCAL(是一种把个体潜在类别参数化的 INDSCAL 的新形式)。也可以对数据构成(Procrustean 转轴)进行比较,并且数据范围也可能扩大到七维数据。

卡罗尔和亚拉伯(Carroll & Arabie, 1980)对 MDS 的各种变式进行了广泛的评述,同时访问相关网站也可以了解到三维应用的文献目录。

应用

对于基本模型以及聚合度量(aggregate measures)的输入来说,通常对于个案数量没有限制。但是对于二维双模数据和三维数据来说,如果个案数量超过 100 个,那么就很难求解(或者解很难解释)。在这种情况下,更适合采用"伪被试(pseudo-subjects)"(经过理论分析或者聚类分析【Cluster Analysis】之后选择的个体子组)。

应用多维度量方法的学科已从最初的心理学和政治科学扩展到经济学等其他社会科学以及生物科学和化学。

——Anthony P. M. Coxon

(李洋译　高勇校)

参考文献

Carroll, J. D., & Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via a N-way generalisation of Eckart-Young decomposition. *Psychometrika*, *35*, 283-299, 310-319.

Carroll, J. D., & Arabie, P. (1980). Multidimensional scaling. *Annual Review of Psychology*, *31*, 607-649.

Coxon, A. P. M. (1982). *The user's guide to multidimensional scaling*. London: Heinemann Educational.

Kruskal, J. B. (1964). Multidimensional scaling by optimising goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*, 1-27, 115-129.

Ramsay, J.O. (1977). Maximum likelihood estimation in MDS. *Psychometrika*, *42*, 241-266.

Takane, Y., Young, F. W., & DeLeeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, *42*, 7-67.

多维性(Multidimensionality)

单维结构存在于如下情境中:单一的基本维度【Dimension】构成了一组观测的基础。例如,包括对政治候选人进行资质评估时的测度,政治积极性量表【Scale】,或者描述某一种人格特质的测度。前提假定是,存在单一的共同维度,候选人、市民或者人格特质可以按这种维度进行排列。

社会科学中的许多理论构念【Constructs】是非常复杂的。单一、同质、潜在的维度可能不足以表达这些复杂的概念。这种复杂性要求我们对这些变量的性质进行更为精细的解释,每组经验观测背后都有不止一个的潜在维度。这就是多维性。为了在进一步的分析中使用这些概念,其多面性必须在测量【Measurement】中反映出来。从这点上,设定多重维度是一种将复杂概念简化为可测量且可理解的多个方面的定义方式,它并不代表本质现实。它的功用在于对理论构念进行操作化【Operationalizition】,以易于理解。

研究人员与研究维度相关的重要工作是发现多元维度并确定哪些在科学中是重要的。假设多维性存在,是因为理论【Theory】表明多维性应该存在,或者因为我们基于手头的经验观测值中的变异性而预测这些潜在维度存在。假设这些维度之间如何相关,然后使用某种经验技术来检验我们的假设【Hypotheses】。这样,这些概念就有了理论和经验基础。

例如,我们对政治意识形态这个概念进行建模,通过人们对政府干预经济的态度来测量与通过人们对公车和平权行动等种族问题的态度来测量,结果是完全不同的。还可以通过人们对诸如流产或者同性恋权利等社会—文化议题的态度来测量,以进一步对意识形态建模。这些都是意识形态概念中的不同维度。

在维度选择上,一个重要的考量是它们对特定目标的重要性。雅各比(Jacoby, 1991)举了一个例子。地图是物理世界的模型【Model】。地图的类型或者维度取决于制作目的或者目标受众。开汽车旅行的人需要一个标示道路的地图。这通常只包括两个维度——纬度和经度。但是,地形学工程师需要更多维度上的细节,如海拔高度。因而,被认为重要的维度的数目取决于实际背景。

——Edward G. Carmines
James Woods
(李洋译　高勇校)

参考文献

Carmines, E. G., & McIver, J. P. (1981). *Unidimensional scaling* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-024). Beverly Hills, CA: Sage.

Jacoby, W. G. (1991). *Data theory and dimensional analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-078). Newbury Park, CA: Sage.

Nunnally, J. C. (1978). *Psychometric theory.* New York: McGraw-Hill.

Weisberg, H. F. (1974). Dimensionland: An excursion into spaces. *American Journal of Political Science*, *18*, 743-776.

多题项量度(Multi-Item Measures)

题项是心理学量表的基本单位。它是用清晰确凿的词汇表达有关被测特征的一个陈述或问题(Haladyna,1994)。在社会科学中,人们用量表【Scale】来测评人们的社会特征,如态度个性、意见、情感状态、个人需求和居住环境。

题项是可以给出"小分(molecular score)"的一道小题("mini" measure)(Thorndike,1967)。在社会科学中,多题项测量优于单一而直接的问题。其中有两个原因:第一,多题项测量的信度【Reliability】要高于单一问题测量。在单一问题的测量中,随着时间的推移,人们给出一致性答案的可能性较低。许多事物可以影响人们的回答(如情绪、他们那天遇到的具体事情)。面对同一个问题,他们可能今天选择是,明天选择不是。还有可能人们提供的答案是错的,或者不同时间对问题的理解不一样。而多题项测量是用多个问题来测量同一个社会议题,最终的综合得分要基于所有问题。人们在多个题项上同时出现上述错误的可能性较低,而综合得分在不同时间上的一致性更强。因而,多题项测量的信度比单一问题更高。第二,多题项测量的效度【Validity】要高于单一问题测量。许多被测社会特征涵盖广泛,不能简单地通过单一问题来评测。多题项测量必定能涵盖测量特征的更多内容,能够完整地反映其构念领域(construct domain)。

这些问题最好通过实例来进行说明。为了评价人们的工作满意度,单一题项测量如下所示:

我对自己的工作不满意。(1=不同意,2=有点不同意,3=不确定,4=有点同意,5=同意)

人们对这个单一问题的回答可能随着时间的推移产生变化。他们对单一问题的回答可能取决于他们的心情或者他们那天在工作上遇到的具体问题,从而差别很大。此外,人们在阅读或回答时可能会犯错误。例如,他们可能没有注意到词语"不",结果想表达不同意的时候却回答了同意。因而,关于工作满意度的单一题项测量显然是信度不高的。另一个问题是人们对他们工作的感受可能不是这样简单。工作满意度是一个很宽泛的概念,它包括许多方面(如对上级的满意度、对同事的满意度、对工作内容的满意度、对报酬的满意度等)。受访者可能喜欢其中某些方面但不喜欢其他方面。单一题项测量大大简化了人们对工作的感受。

多题项测量可以减少上述问题。从多题项测量中得出的结果会更具一致性。使

用多个题项能通过平均化来消除随机误差(Spector,1992)。也就是说,总共有20个题项,如果一个受访者在1个题项上犯了一个错误,对总体分值的影响也微乎其微。更重要的是,多题项测量能让受访者对他们工作的不同方面进行描述。这将会极大地提高测量的精度和效度。因而,多题项测量是社会科学研究中最重要、最经常使用的工具。

——Cong Liu

(李洋译　高勇校)

*也可参见度量【Scaling】。

参考文献

Haladyna, T. M. (1994). *Developing and validating multiplechoice test items.* Hillsdale, NJ: Lawrence Erlbaum.

Spector, P. (1992). *Summated rating scale construction: An introduction.* Newbury Park, CA: Sage.

Thorndike, R. L. (1967). The analysis and selection of test items. In D. Jackson & S. Messick (Eds.), *Problems in human assessment* (pp. 201-216). New York: McGraw-Hill.

多层次分析(Multilevel Analysis)

大部分统计推论都是基于对某一类分析单位【Unit of Analysis】(如个人样本、国家样本或学校样本)进行重复观测的基础上。对这些观测的分析通常基于如下假设:抽中的分析单位本身或统计模型中的相应残差是独立同分布的。但是,社会现实的复杂性和社会科学理论常常要求更为复杂的数据集【Complex Data Sets】,其中包括多种分析单元。例如,要研究教育成就,学生、教师、班级、学校都是重要的分析单位;要研究组织,雇员、部门和公司都是分析单元;要进行跨国比较,个体和国家(或地区)都是分析单位;在概化理论【Generalizability Theory】研究中,每个因素都定义了一类分析单位;在元分析【Meta-Analysis】中,所收集到的不同研究、进行研究的不同群体、研究中的不同被试或被访者都是不同分析单位,也是未解释变差的来源。不同类型的分析单位经常是(但并不总是)层级嵌套的(如学生嵌套于班级,班级嵌套于学校)。多层次分析这一术语是指适用于分析由多种类型的分析单位组成的数据集的统计方法。多层次分析中的层次是不同类型的分析单位的另一个称谓。每个分析层次都对应于一个总体【Population】,因此,多层次研究涉及多个总体。在第一个例子中,就有4个总体:学生、教师、班级和学校。在一个严格嵌套的数据结构中,最具体的层次被称为第一层或者最低层次。例如,在一个学生嵌套于班级、班级嵌套于学校的数据集中,学生、班级和学校分别是第一层次、第二层次和第三层次。

层级线性模型

多层次分析中一种最重要的方法是回归【Regression】分析的一种变体,用来分析层级嵌套数据。主模型是层级线性模型【Hierarchical Linear Model,HLM】,它是一般线性模型【General Linear Model】的扩展,其误差或残差的概率模型结构体现出了数据的层级结构。由于这个原因,多层次分析常常也被称为层级线性模型。举例来说,研究者要研究大学生毕业后的年收入多大程度上取决于大学时期的学业成绩。我们假设

研究人员从一些高校中收集到了数据，比方说，从高校总体中抽取到了三十多所作为样本，然后再从每所高校中对 15 年前毕业的学生进行随机抽样。对每个学生都了解了他们的当前收入（变量 Y）和在校平均成绩（变量 X，其中 $X=0$ 表示最低及格分数）。毕业生用字母 i 表示，高校用字母 j 表示。由于毕业生嵌套在高校中，每所高校的毕业生 i 都分别从 1 开始进行编号，变量表示为 Y_{ij} 和 X_{ij}。设高校为 j，就可以在下列模型的基础上开始分析：

$$Y_{ij} = a_j + b_j X_{ij} + E_{ij}$$

这只是一个线性回归模型，其中截距【Intercepts】a_j 和回归系数 b_j 取决于高校，因此，下标为 j。高校是来自总体的随机样本，这一事实反映在截距项 a_j 和回归系数 b_j 的随机变差上。设截距（在所有高校的总体中）的均值为 a，具体高校的离差为 U_{0j}，那么 $a_j = a + U_{0j}$。同样可以将回归系数分解为总体均值和具体高校离差 $b_j = b + U_{1j}$。用这些等式进行替换就得到

$$Y_{ij} = a + bX_{ij} + U_{0j} + U_{1j}X_{ij} + E_{ij}$$

这个模型有 3 种不同类型的残差项：第一层残差项 E_{ij}，第二层残差项 U_{0j} 和 U_{1j}。第一层残差项因毕业生不同而有所不同；第二层残差项因高校不同而有所不同。可以将残差项解释如下：对于 U_{0j} 值高的高校来说，其具有最低及格成绩 $X=0$ 的毕业生有相对较高的期望收入，即 $a + U_{0j}$；对于 U_{1j} 值较高的高校来说，个体 GPA 对毕业生期望收入的影响更大，即 $b + U_{1j}$；给定毕业生的毕业高校 j 和 GPA 成绩 X_{ij}，那么，E_{ij} 值较高的毕业生收入相对较高。

这个方程就是 HLM 的一个实例，在其一般形式中，模型可以包括不止一个自变量。方程第一部分 $a + bX_{ij}$ 是模型的固定部分，与线性回归分析中一样，它是自变量的一个线性函数。第二部分 $U_{0j} + U_{1j}X_{ij} + E_{ij}$ 是方程的随机部分，它比线性回归分析中的随机残差更复杂，因为它不仅包括毕业生之间的未解释变差 E_{ij}，还包括高校之间的未解释变差 $U_{0j}+U_{1j}X_{ij}$。模型的随机部分就是多层线性模型和一般线性模型的不同之处。可以对一个双水平模型的随机部分进行最为简化的设定，就是设定在第二层单位上截距有所不同但回归系数相同。这被称为随机截距模型，在本例中即

$$Y_{ij} = a + bX_{ij} + U_{0j} + E_{ij}$$

在第二层单元上回归系数也会随机变化的模型被称为随机斜率模型（在回归图上回归系数就是回归线的斜率）。

HLM 中的因变量 Y 一定是在最低（即最具体的）层次上。HLM 模型的一个重要特点是自变量或者解释变量可以在任何分析层次上。例如，在对高校毕业生收入的研究中，研究人员对以高校排名测量的高校质量对毕业生收入的影响感兴趣，而且可以获得高校排名得分 Z_j。在之前的模型中，高校分析层次的残差项 U_{0j} 和 U_{1j} 反映了高校之间未解释的变异。这个变异就可以用高校层次的变量 Z_j 来部分地解释，方程如下：

$$a_j = a + c_0 Z_j + U_{0j}, b_j = b + c_1 Z_j + U_{1j}$$

这个方程可以看成 a_j 和 b_j 两个量在第二层的一个线性回归方程，这两个量难以进行直接观察。将这些方程代入第一层方程 $Y_{ij} = a_j + b_j X_{ij} + E_{ij}$ 中，就可得到新的模型：

$$Y_{ij} = a + bX_{ij} + c_0 Z_j + c_1 X_{ij} Z_j + U_{0j} + U_{1j}X_{ij} + E_{ij}$$

与之前的模型相比，现在的参数 a 和 b 以及残差项 E_{ij}，U_{0j} 和 U_{1j} 都有了不同的意义。

这个模型的固定部分得到了扩展，但是随机部分保留了原来的结构。固定部分中的 $c_1 X_{ij} Z_j$ 项是第一层变量 X 和第二层变量 Z 之间的交互效应【Interaction Effect】。回归系数 c_1 表示高校情况（Z）能够在多大程度上改变个人成绩（X）对收入（Y）的影响；这样的效应被称为跨层次交互效应。多层次模型能够广受欢迎的一个重要原因，就在于从这个模型中能够看出高校情况［“宏观层次（macro level）”］是如何对个体层次［“微观层次（micro level）”］变量之间的关系产生影响的（见 DiPrete & Forristal，1994）。

有一个参数能够描述数据中两个层次的相对重要性，这个参数就是组内相关【Intraclass Correlation】系数，在相应词条和方差分量模型【Variance Component Model】词条中对此有所介绍。类似的方差比，当用于残差（即未解释的）方差时，被称为组内残差相关系数（residual intraclass correlation coefficient）。

假设、估计和检验

HLM 的标准假设包括：模型如方程所示是线性的，残差服从正态分布，各个层次的残差彼此独立，同一层次上不同单元的残差彼此独立。但是，同一单元的不同残差（如上面模型中的随机截距 U_{0j} 和随机斜率 U_{1j}）可以相关；它们被假定服从多元正态分布。在这些假设前提下，上面实例中的 HLM 意味着同一所高校毕业生之间存在相关，因为他们受到了同一所高校的影响。从技术角度来看，这是由于他们关于 Y_{ij} 的方程中高校层次的残差 U_{0j} 和 U_{1j} 是相同的。这种不同个案之间的相依关系，与回归分析中更为传统的一般线性模型的假设大为不同。

HLM 的参数可以通过最大似然【Maximum Likelihood】方法进行估计。各种算法【Algorithm】在 1980 年代得到了长足发展（参见 Goldstein，2003；Longford，1993）；一个重要的算法是迭代再加权最小二乘算法（iterative reweighted least squares algorithm）（见广义最小二乘【Generalized Least Squares】的词条），这种方法交替性地对固定部分的回归系数和随机部分的参数进行估计。这些模型的回归系数可以通过 t 检验【t-Tests】或者 Wald 检验进行检验，而决定随机部分结构的参数则可通过似然比检验【Likelihood Ratio Test】（也称作偏差检验）或者卡方检验来进行检验。从 1980 年代起，这些方法已经可以在多层次分析专用软件（如 HLM 和 MLwiN）中实现，随后一些统计软件包（如 M-Plus）将多层次分析纳入一个更广义的统计方法集合；此外，在一些常规的统计包（如 SAS 和 SPSS）中也有这种分析方法。戈尔茨坦（Goldstein，2003）介绍了各种软件的用途。

多层次

正如上面例子所表明的，在实际调查中经常会涉及两个以上的分析层次。在教育研究中，学生和教师通常对成绩的影响最大，但是班级中学生群体营造的群体环境、由学校营造的组织环境、由邻里所界定的社会环境同样会产生重要影响。在一项关于学生学业成绩的研究中，上述分析层次上所定义的各个变量都是解释性变量。如果某个分析层次确有影响，那么我们会想到这个影响不会完全被该分析层次上的变量所解释，这个层次的分析单位之间会有一些未解释变差。通过将未解释变差作为随机残差变异纳入模型，可以反映出这一点。第一种类型的残差变异是这个层次的单元的随机主效应，例如，上面例子中双层次模型的随机截距 U_{0j}。此外，数值变量（比如学生层次的变量）的效应也可能在所考察层次上的不同单元之间有所差异，这一点可以通过上面提到的 U_{1j} 等随机斜率来纳入模型中。采用多层次分析所得出的一类重要结论，就是在

不同层次上对未解释的变异进行分解。这在不含随机斜率的方差成分模型【Variance Component Model】的词条中有所讨论。每个层次上有多少未解释变异,这为研究人员指明了寻求进一步解释的方向。

不同分析层次之间可能是嵌套的或者交叉的。如果较高层次的单元就是较低层次单元的集合(即较低层次的每个单元都被完全包含在某个较高层次单元中),那么,可以说一个分析层次(较低层次)嵌套在另一个较高层次中。否则,我们就说这些层次之间是交叉的。与嵌套层次相比,交叉层次分析更容易在分析中遇到困难:算法在进行估计时可能会产生更多的收敛问题,在不同层次上分解变异的经验结论可能并不清晰,同时在概念和理论建模中也存在更多的模糊性。

使用多层次分析模型需要足够丰富的数据才能进行统计分析。需注意的是,对于每个分析层次来说,数据中的单元构成了一个来自相应总体的样本。尽管任何经验法则都有其局限性,但是一般来说,如果样本量低于20(即一个分析层次的单元数小于20个),样本中包括的总体(即这个分析层次)信息就相当有限,而如果样本量小于10,结论就大可怀疑。

纵贯数据

在纵贯研究【Longitudinal Research】中,HLM也大有用武之地。在最简单的纵贯数据结构中,个体被重复测量多次,重复测量就构成了较低(第一)层次,个体构成了较高(第二)层次。多数纵贯研究会有一个有意义的数值型时间变量:如在一个实验研究中,可能是从实验开始起过了多长时间;在一个发展研究中,可能是年龄。尤其是在非平衡的纵贯数据结构中,不同个体的观测数量和次数是不同的,此时建构多层次模型将是水到渠成的一种方便做法。结果变量与时间维度之间的相依关系是模型中很重要的方面。通常,刚开始可以用线性相依来试探。这相当于把时间测量作为一个解释性变量,这个变量的随机斜率表示不同个体有不同的变化率(或增长率)。但是,与时间相依的模型通常是非线性的。在一些情况中,可以对时间进行非线性转换(例如多项式或者样条曲线),然后假定时间变量转换后模型就是线性的,这样就可以继续用HLM进行非线性问题建模(见Snijders & Bosker,1999,Chap.12)。在其他情况中,最好不要采用相对简单的线性模型,而是建构非线性模型,或者第一层残差存在自相关的模型(参考Verbeke & Molenberghs,2000)。

非线性模型

残差项服从正态分布的假设并不总是能够成立,虽然有时通过因变量转换会使这一假设变得更现实一些。尤其是如果因变量是二分变量或离散变量,就需要采用其他模型。正如广义线性模型【Generalized Linear Model】是一般线性回归模型的一种扩展,对HLM进行非线性扩展就可以使Logistic回归【Logistic Regression】和Logit模型【Logit Model】等模型扩展到多层次的情形中。这些模型被称为层次广义线性模型(hierarchical generalized linear models)或广义线性混合模型(见词条层次非线性模型【Hierarchical Nonlinear Models】)。

——Tom A. B. Snijders

(李洋译　高勇校)

参考文献

de Leeuw, J., & Kreft, I. G. G. (Eds.). (in press). *Handbook of quantitative multilevel analysis.* Dordrecht, The Netherlands: Kluwer Academic.

DiPrete, T. A., & Forristal, J. D. (1994). Multilevel models: Methods and substance. *Annual Review of Sociology*, *20*, 331-357.

Goldstein, H. (2003). *Multilevel statistical models*

(3rd ed.). London: Hodder Arnold.
Hox, J. J. (2002). *Multilevel analysis, techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.
Longford, N. T. (1993). *Random coefficient models*. New York: Oxford University Press.
Raudenbush, S. W., & Bryk, A. S. (2001). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.
Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York: Springer.

多方法多特质研究(Multimethod-Multitrait Research)

参见多方法研究【Multimethod Research】。

多方法研究(Multimethod Research)

多方法研究就是用两种及以上的数据或方法来研究某个问题或者某几个高度关联的问题。这种研究通常也被称为混合方法。混合方法研究(mixed-method research)的道理在于,大多数社会研究都是基于某一种研究方法得到的研究结果,但是如果测量误差会影响研究结果,那么单项研究中得到的结论和推论就可能有误。单一方法几乎不可能估计出测量误差【Measurement Error】对研究结果有多大影响,因此,单一方法研究总是在这方面备受质疑。

多方法研究和测量

多方法研究的基本原理是基于三角测量法【Triangulation】的。三角测量法认为研究人员要确保自己没有过度依赖一种研究方法,在研究某问题时应当采取多种测量程序。因此,多方法研究的根据就在于它能够增强人们对研究结果的信心,这也在很大程度上解释了它日渐普及的原因。

在不同的测量情境中,混合方法研究也有不同的内涵。第一种情形是要研究的构念(construct)有不同的测量方式(如有不同方式可以测量工作满意度),这就需要将两种及以上的测量方法结合起来使用。第二种情形是有两种及以上的数据收集方法。例如,要对公司中工种的特性进行考察,同时用结构化观察法和结构化访谈法来对同一被试调查各个工种的特性,然后评估两组数据之间的一致性水平。

混合方法多特质矩阵

混合方法常被提及的一种情形是坎贝尔和菲斯克(Campbell & Fiske, 1959)提出的混合方法多特质矩阵(mixed-method-multitrait matrix)。不妨这样说,我们有一个构念,这个构念在概念上有5个维度,或像坎贝尔和菲斯克所称的5个“特质”。我们可以发展出子量表来测量每个维度,同时兼用结构化访谈法和结构化观察法调查每个

维度。我们能够生成一个矩阵来汇总不同特质间的方法内相关系数（within-method correlations）和方法间相关系数（between-method correlations）。矩阵能解决两个问题：第一，它可以对测量进行判别效度【Discriminant Validity】检验，因为方法内相关不是太大。第二，它可以对不同方法得到的分值的收敛效度进行检验，这一点在此尤为重要。得出的相关性越大，收敛效度越大，研究结果的可信度就越高。表 1 显示了最终得出的矩阵：

表 1　混合方法多特质矩阵

访谈	访谈					观测				
	T_1	T_2	T_3	T_4	T_5	T_1	T_2	T_3	T_4	T_5
T_1	—									
T_2	$r_{i2,i1}$	—								
T_3	$r_{i3,i1}$	$r_{i3,i2}$	—							
T_4	$r_{i4,i1}$	$r_{i4,i2}$	$r_{i4,i3}$	—						
T_5	$r_{i5,i1}$	$r_{i5,i2}$	$r_{i5,i3}$	$r_{i5,i4}$	—					
观测										
T_1	$\boldsymbol{r}_{i1,o1}$					—				
T_2	$r_{i2,o1}$	$\boldsymbol{r}_{i2,o2}$				$r_{o2,o1}$	—			
T_3	$r_{i3,o1}$	$r_{i3,o2}$	$\boldsymbol{r}_{i3,o3}$			$r_{o3,o1}$	$r_{o3,o2}$	—		
T_4	$r_{i4,o1}$	$r_{i4,o2}$	$r_{i4,o3}$	$\boldsymbol{r}_{i4,o4}$		$r_{o4,o1}$	$r_{o4,o2}$	$r_{o4,o3}$	—	
T_5	$r_{i5,o1}$	$r_{i5,o2}$	$r_{i5,o3}$	$r_{i5,o4}$	$\boldsymbol{r}_{i5,o5}$	$r_{o5,o1}$	$r_{o5,o2}$	$r_{o5,o3}$	$r_{o5,o4}$	—

在该矩阵中，$r_{i4,o2}$ 表示用结构化访谈（i）测量的特质 T_4 和用结构化观测（o）测量的 T_2 之间的相关性。5 个粗体字显示的相关系数（如 $\boldsymbol{r}_{i1,o1}$）是收敛效度相关性，表示对每个特质进行测量的两种方法在多大程度上是一致的。

在混合方法研究中即使收敛效度很高，也仍然需要小心。我们很容易假设只要有收敛效度，就能保证研究的测量效度。但是，对构念或研究问题进行测量的两种方法有可能都有问题。因此，虽然收敛效度能够增强我们的信心，但是并不能完全排除误差仍然存在的可能性。

定量研究和定性研究的结合

到现在为止，对混合方法研究的论述还是基于测量角度，尤其是基于三角测量法传统上进行的。但是，对混合方法研究的讨论已经日益延伸到定性数据和定量数据的收集上来。换句话说，混合方法研究逐渐开始涉及定量研究和定性研究的结合。从这个角度来看，对混合方法研究和三角测量法的讨论，不仅是一个测量问题，同样也涉及不同的数据收集方式。

定量研究【Quantitative Research】和定性研究【Qualitative Research】有时被看成彼此不兼容的不同研究范式【Paradigms】。许多人认为定量研究和定性研究来源于完全不同的认识论和本体论传统。他们认为,尽管可以在一项调查研究中同时采用结构化访谈(这是典型的定量研究方法)和民族志【Ethnography】(这是典型的定性研究方法),但是这并不是两种方法的真正结合,两种方法也不可能结合,因为两种方法有着不同的基础原则。这些学者认为,这种结合不过是把两种互不兼容的方法肤浅地组合在一起而已。

尽管如此,大多数研究人员持有一种更为实用主义的立场,他们也承认定量研究和定性研究代表了不同的认识论和本体论准则(如实证主义和解释主义【Positivism *vs.* Interpretivism】以及客体主义和建构主义【Objectivism *vs.* Constructionism】),但是他们认为把这两种方法的优点结合起来大有好处。定量研究和定性研究之间表面上的不可通约性(incommensurability)是可以解决的,那些认识论和本体论议题大可忽略,研究方法和数据来源并没有像我们想的那样和认识论预设密切关联。根据这种看法,研究方法只是一种收集资源的技术,它和有效知识的本质这类宏大的思考不甚相干。

有人会认为,把定量研究和定性研究组合起来并不是一种如多策略研究【Multi-Strategy Research】那种的混合方法(Bryman,2001)。这种倾向反映了一种观点,认为定量研究和定性研究是在社会研究方法上的不同研究策略,每种方法都有自己的一整套研究方法和研究设计。这种区分表明,可以通过交叉分类形成4种分类,如图2所示。

与混合方法多特质矩阵相关联的混合方法研究本质上属于表2中的单元格1,因为这种研究是把与定量研究相联系的两种数据收集方法结合起来。单元格4中的研究是把与定性研究相关联的两种及以上方法结合起来。事实上,许多民族志研究几乎天生就是一种混合方法研究,因为民族志学者通常不只是在参与观察。他们有多种渠道来辅助自己进行观察,如访谈关键知情人、查阅档案、进行非参与观察。

表2　混合方法研究的单策略和多策略方法

	定量研究方法	定性研究方法
定量研究策略	单策略多方法研究1	多策略多方法研究2
定性研究策略	多策略多方法研究3	单策略多方法研究4

定量和定性研究相结合的方法

单元格2和单元格3的相同之处在于,它们都包括多策略混合方法研究,在这种研究中,至少采用了一种与定量研究和定性研究都有关系的研究方法或者数据来源。通过考察两种研究策略相结合的不同方式,可以进一步深入地分析多策略混合方法,多位作者已经对这一问题进行了论述。摩根(Morgan,1998)提出了根据两个原则对多策略研究进行分类的机制:在定量研究和定性研究中何者是收集数据的主要方法,何种研究方法先于另一种方法进行。这样的区分交叉组合,就构成了混合方法研究(楷体字表示主要方法)的4种类别:

(1)定性→定量。例如,研究人员为了完善多题项量表的题项而实施半结构化访谈;或者进行探索性的定性研究来生成假设以便随后用定量方法进行检验。

(2)定量→定性。例如,先进行一项问卷调查,基于调查中表明的特征选择特定的个人,然后用定性方法进一步进行重点研究。

(3)定量→定性。例如,研究人员会开展一些半结构化访谈或者参与观察,目的是弄清楚有哪些因素可以用来解释先前问卷调查中发现的变量之间的关系。

(4)定性→定量。例如,在民族志研究中发现了变量间存在有趣的关系,然后开展一项问卷调查来确定这种关系在多大程度上有外部效度【External Validity】。

这4种类型的多策略研究就是定量和定性研究的不同结合方法,是对表2的单元格2和单元格3的进一步分析。摩根(Morgan,1998)的分类并没有穷尽,因为他没有考虑如下情形:(a)定量和定性研究同步开展的多策略研究;(b)资料收集没有主次之分的多策略研究(即它们的分量是相同的)。克雷斯韦尔(Creswell,1995)的分类中包括了这两种多策略研究:

(5)平行/同步研究。定量和定性研究方法基本上是同步实施的。

(6)两种方法同等重要的研究。两个研究策略在整个研究计划中地位相同。

这两种类型经常同时发生,因为在混合方法研究中要用不同研究方法来考察所研究的现象的不同方面。支持使用多策略研究的一种更常见的观点认为,只使用定量或定性研究不足以触摸到研究问题的所有维度。在这种情形下,定量和定性研究方法可能或多或少是同步进行的,有着同样的重要性。

塔沙可利和特德利(Tashakkori & Teddlie,1998)更进一步指出了第七种多策略研究的类型:

(7)不同层次使用不同方法的设计。研究人员"在数据的不同层次使用不同类型的方法"(p.18)。这种研究的例子是,研究人员在不同的组织层次上收集不同类型的数据(对组织成员进行问卷调查,对部门领导进行深度访谈,对工作实践进行观察)。

此外,塔沙可利和特德利(Tashakkori & Teddlie,1998)认为,我们现在谈到的这类多策略研究就是他们所称的"混合方法"研究。他们认为应区分混合方法研究与混合模型设计,混合模型设计是把定量和定性研究的不同阶段结合起来。例如,研究人员以一种本质上探索性的方式收集了定性资源,然后对这些数据进行定量分析而非定性分析,这就是混合模型研究。

进行多策略研究最常提及的理由就是试图确立不同研究发现的收敛有效性;换句话说,这种理由诉诸三角测量法的逻辑。如果定量结果可以用定性证据来证实(反过来也一样),研究的可信性就增强了。但是,这种想法也有问题。第一,定量和定性研究来自截然不同的认识论和本体论立场,很难说某一组证据能够对另一方形成支持或否定。第二,和所用三角测量法的情形一样,很难处理两组证据之间的不一致。一个常见的错误是认为某一组比另一组更有效,进而用前者作为效度的标准。第三,人们通常认为三角测量法有较强的朴素唯实论【Realism】色彩,而许多定性研究者认同的是建构论(constructionist),他们否认社会世界中绝对有效知识的可能性,因此,定性研究实践与三角测量法有不一致之处。第四,一般很容易假定多策略研究必然优于单一策略研究,但是,如果总体研究设计不允许一种研究策略能够在其他研究策略的基础上有所增益,就并非如此了。第五,有些社会研究者对各种研究方法或方式都有所了解,但是有些研究者只精通某一种研究方法,虽然这也表明

了团队合作很重要。

三角测量的理念总是假定研究是有计划的。但是采用前面提到的7种多策略研究方法做研究时,可以想见所得到的定量和定性数据可能与某一具体议题有关,结果构成了一个未计划的三角测量研究。当从两组证据中发现收敛有效时,研究人员对比并不会感觉稀奇,但是如果两组数据并未收敛,研究人员就需要进一步思考导致结果相抵触的深层次原因。迪肯、布里曼和芬顿(Deacon, Bryman, & Fenton, 1998)对于社会研究成果对大众媒体上的展现进行多策略研究,他们发现定量研究和定性研究的结果有诸多差异。这些差异并非源于有计划的三角测量。例如,他们发现,对社会科学家的问卷调查结果表明他们对媒体的展现感到满意,但是当就媒体对他们研究的报道进行深度访谈时,他们变得批判性更强。这种不一致促使研究者考虑提问的不同类型和不同情境,以及这对于理解社会科学中媒体的作用有何意义。

毋庸置疑,混合方法研究在社会研究中的应用越来越多。它的吸引力在于它可能增强调查的效度。混合方法研究可以涉及相同的研究策略或不同的研究策略,因此,是一种处理研究问题的灵活方式,虽然它可能更加费时费力。混合方法研究尤其是多策略研究会出现阐释问题,尤其是研究发现存在不一致时,但是研究者对于此类问题未雨绸缪将会增强他们研究工作的可信度。

——Alan Bryman

(李洋译　高勇校)

参考文献

Bryman, A. (2001). *Social research methods*. Oxford, UK: Oxford University Press.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.

Creswell, J. W. (1995). *Research design: Quantitative and qualitative approaches*. Thousand Oaks, CA: Sage.

Deacon, D., Bryman, A., & Fenton, N. (1998). Collision or collusion? A discussion of the unplanned triangulation of quantitative and qualitative research methods. *International Journal of Social Research Methodology*, *1*, 47-63.

Morgan, D. L. (1998). Practical strategies for combining qualitative and quantitative methods: Applications for health research. *Qualitative Health Research*, *8*, 362-376.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.

多项式分布(Multinomial Distribution)

假设我们有一个 N 个被试(或 N 个个体,或 N 个题项)组成的随机样本【Random Sample】,每个被试都属于 k 个类别之一,这些类别互不重叠。也就是说,每个被试能且只能被划入一个类别。p_i 是一个被试属于类别 i 的概率,$i=1,2,\cdots,k$,其中 $0 \leqslant p_i \leqslant 1$ 且 $\sum p_i = 1$。这些类别被称为是互斥且穷尽的。

假设随机变量【Random Variable】X_i 是样本中属于类别 i 的被试数目,$i=1,2,\cdots,k$。很显然,$\sum X_i = N$。此时,k 个随机变量的联合概率分布【Probability Distribution】被称为多项式分布,即

$$f(X_1,X_2,\cdots,X_k)$$
$$=\left(\frac{N}{X_1,X_2,\cdots,X_k}\right)p_1^{X_1}p_2^{X_2}\cdots p_k^{X_k} \quad (1)$$
$$=\frac{N!}{X_1!\ X_2!\ \cdots X_k!}p_1^{X_1}p_2^{X_2}\cdots p_k^{X_k} \quad (2)$$

其中的参数【Parameter】为 $N,p_1,p_2,\cdots,p_k$。分布中总共只有 $k-1$ 个随机变量,因为,加总是一个定值, $\sum X_i=N$。每个 X_i的均值和方差分别为 Np_i和 $Np_i(1-p_i)$,X_i和 X_j的协方差是$-Np_ip_j$。每个 X_i的边缘分布是含参数 N 和 p_i 的二项式分布【Binomial Distribution】,参数是 N 和 p_i。

举例说明,假设我们旋转一个正常骰子10次,想看看 1 和 6 出现一次、其他 4 个数字出现两次的概率,也就是说,$P(X_1=X_6=1,X_2=X_3=X_4=X_5=2)$。对于正常骰子来说每个 $p_i=1/6$, 因此计算结果是 0.003 751。

(1)中的大括弧的项,或是(2)中包含阶乘的项,被称为多项式系数。

如果 $k=2$,多项式分布就简化为常见的含参数 N 和 p_1 的二元分布,其中的两个类别通常被称为成功和失败,X_1是成功数,$X_2=N-X_1$是失败数。

——Jean D. Gibbons

(李洋译 高勇校)

多项 Logit(Multinomial Logit)

多项 Logit(MNL)是一种统计模型,用来研究因变量是定类的且有着 3 个及以上类别的问题。MNL 对每个观测值属于因变量不同类别的概率【Probability】进行估计。MNL 将个体观测特征作为自变量,自变量的效应可以在因变量的不同类别中有差异。这不同于类似的条件 Logit【Conditional Logit】模型,后者将选择类别(choice categories)的特征作为自变量,因而自变量的效应在各个选择类别上都是相同的。事实上,模型中也可以同时包含选项的特征和个体的特征,这通常也被称作条件 Logit 模型。

有多种不同的方法可以得到 MNL 模型。在社会科学中最常用的方法是将 MNL 作为一种离散【Discrete】选择模型。在每个观测中,一位行动者从包括 J 个选项中选择一项,其中 J 是因变量中的类别数。行动者 i 从选项组中选择选项 k 的效用是

$$U_{ik}=X_i\beta_k+\varepsilon_{ik}$$

y 是包含 J 个无序类别的因变量。因而,行动者 i 选择选项 k 而非 j 的概率是

$$\begin{aligned}\Pr(y_i=k)&=\Pr(U_{ik}>U_{ij})\\&=\Pr(X_i\beta_k+\varepsilon_{ik}>X_i\beta_j+\varepsilon_{ij})\\&=\Pr(\varepsilon_{ik}-\varepsilon_{ij}>X_i\beta_j-X_i\beta_k)\end{aligned}$$

麦克法登(McFadden,1973)证明,如果 ε 在所有的观测和选项上都是独立同分布的且服从耿贝尔分布(Gumbel distribution),由上述离散选择模型就可得到 MNL 模型。MNL 对 $\Pr(y_i=k|x_i)$进行估计,其中 k 是因变量的 J 个类别之一。MNL 模型可以表示为

$$\Pr(y_i=k\mid x_i)=\frac{e^{x_i\beta_k}}{\sum_{j=1}^{J}e^{x_i\beta_k}}$$

MNL 采用最大似然【Maximum Likelihood】技术进行估计。为了对模型进行识别,必须对 β 增加限制条件。通常某一

个类别的β被限定为等于0,但也可以采用其他限制条件(例如,$\sum_{j=1}^{J}\beta_j=0$)。如果某一类别的β被限定为0,那么这个类别就是基准类别,这时所有其他的β_j表示自变量对选择类别J而非选择基准类别的概率的影响。MNL的系数不能解释为回归系数【Regression Coefficients】,要判定自变量对因变量的效应必须使用其他方法[如考察当自变量变化时$\Pr(y_i=k)$的变化]。

表1 职业选择的MNL估计(体力劳动的系数标准化为0)

自变量	系 数	标准误
	S/M 系数	
教育程度	0.76*	0.18
年龄	−0.09	0.10
女性	0.17	0.15
常数	0.51	0.42
	P/M 系数	
受教育程度	0.89*	0.16
年龄	0.21	0.15
女性	0.27*	0.13
常数	−0.03	0.67
观测值数量	1 000	

注:M=体力工人;S=技术人士;P=专业劳动。统计显著性水平为0.05。

对于每个因变量类别的成对比较来说,MNL在逻辑上等同于估计一系列二分Logit模型对因变量中的类别进行两两比较,但是MNL更有效率,因为它能同时使用所有的数据。MNL要求选择过程具有无关选项独立性(Independence of Irrelevant Alternatives,IIA)[③],但这在很多情形下可能并不成立。

作为MNL的一个应用案例,可以考察下面的问题,即对年龄、性别和教育程度对职业选择的效应进行估计。将职业分成3个类别:体力劳工(M)、技术工人(S)和专业人士(P)。这些选择的MNL估计结果见表1。

需要注意的是,系数会因选项不同而有所不同,其中一个选项的系数已经被设定为0。MNL模型的结果表明,随着受教育程度的提高,人们会更有可能成为技术工人和专业人士而不是体力劳工。女性更有可能选择成为专业人士而非体力劳工。

——Garrett Glasgow

(李洋译 高勇校)

参考文献

Long, J. S. (1997). *Regression models for categorical and limited dependent variables.* Thousand Oaks, CA: Sage.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.),

③ "无关选项独立性"是指如果行动者从一个集合T中选择了选项x,而x又是集合T中一个子集S的成员,那么,行动者也必然会在子集S中选择选项x。换言之,那些未被选择的选项被剔除后,x仍然是最佳选项。——译者注

Frontiers of econometrics (pp.105-142). New York: Academic Press.

Train, K. (1986). *Qualitative choice analysis: Theory, econometrics, and an application to automobile demand.* Cambridge: MIT Press.

多项 Probit(Multinomial Probit)

回归【Regression】分析假定了因变量【Dependent Variable】是连续变量(或者至少是定距层次)。对于社会科学中的许多研究问题来说,连续变量【Continuous Variable】是无法观测到的,只能通过二分变量【Dichotomous Variable】来显现。例如,无法观测的变量可能是选民选择 A 候选人而非 B 候选人的概率,而可观察变量则是在 A 和 B 之间的实际选择。如果因变量是二分变量,变量连续性的假设就不再成立,分析人员必须选择其他估计方法,比如 Probit 和 Logit 方法。

有时无法观测的连续因变量显现为有着 3 个及以上取值的定类变量或者定序变量。例如,被调查者经常被要求从非常不同意到非常同意的五分量表中进行选择,或者议员可能会面临项目资金应该削减、保持不变、增加的三重选项。测量此类数据的变量可称为多项变量、多元变量或者 n 分变量。多项 Probit 和多项 Logit【Multinomial Logit】是标准 Probit 和 Logit 的一种变换形式,它们可用来分析多元的因变量。

二项 Probit 的本质是在某一观测中自变量【Independent Variables】取值给定时,估计因变量取值为 0 或 1 的概率【Probability】。对观测值进行分类就是要确定哪个值的概率估计更高。在多项 Probit 中,不仅要估计自变量系数【Coefficients】,同时还要估计阈值参数(threshold parameters),阈值参数一般用 μ_j 表示。要记住的是,所需要的阈值参数的数量要比因变量的类别数少 1 个,如果有 3 个类别,只有两个阈值参数需要进行估计,分类过程如下:

$$y = 1 \quad \text{如果 } y^* \leqslant \mu_1,$$
$$y = 2 \quad \text{如果 } \mu_1 \leqslant y^* \leqslant \mu_2,$$
$$y = 3 \quad \text{如果 } \mu_2 \leqslant y^*$$

其中,y 是观测到的因变量,y^* 是不可观测变量。由于操作原因,μ_1 一般需要设定为 0。因而要估计的参数数目比类别数目少 2 个。

对于数据组中的每个观测,都可以估计出它选择因变量中某一个类别的概率。一般而言,一个观测选择某一类别的概率可以这样计算出来:

$$\text{Prob}(y = j) = \Phi\left(\mu_j - \sum_{i=0}^{n} \beta_i x_i\right) - \Phi\left(\mu_{j-1} - \sum_{i=0}^{n} \beta_i x_i\right)$$

其中,j 代表某一具体类别,μ_j是相应类别的阈值参数,$\beta_i x_i$ 是估计系数与相应的 n 个自变量的取值的乘积($\beta_0 x_0$是常数项),Φ 是累积正态概率分布。注意,所有概率都应大于或等于 0,必须满足条件 $\mu_1 < \mu_2 < \mu_3 < \cdots < \mu_j$(记得 μ_1 通常被设定为 0)。考虑概率分布的累积性,上式可以简化为 $\text{Prob}(y = j) = \text{Prob}(y \leqslant j) - \text{Prob}(y \leqslant j - 1)$。

μ_j 的显著性可以用通常的 t 检验【t-Test】来检验。如果 μ_j 估计显著且为正,就说明因变量的类别是定序的。

表 1 是一个多项 Probit 的结果示例。

因变量Y编码为1,2或3。3个自变量X_i编码都是0或1。将一个阈值参数设定为0之后,三分因变量仅剩一个阈值参数需要估计。估计出的阈值参数不仅统计显著【Statistically Significant】,而且取值为正,大于那个被设定为0的参数。

表1　多项 Probit 估计的样本结果

因变量 Y		
自变量	估计系数	标准误
X_1	0.90	0.29
X_2	0.55	0.26
X_3	−0.31	0.31
常数	0.58	—
$\hat{\mu}$	0.97	0.15
N	107	
−2 对数似然比 $df=3$	96.23	

为了计算给定观测值落入因变量每个响应类别的概率,设$X_1=1, X_2=1$且$X_3=1$。根据上面的公式,可以使用累积正态概率分布计算得到

$$\text{Prob}(y=1)=\Phi\left(-\sum_{i=0}^{3}\beta_i x_i\right)=0.041\ 8$$

$$\text{Prob}(y=2)=\Phi\left(\mu-\sum_{i=0}^{3}\beta_i x_i\right)-\Phi\left(-\sum_{i=0}^{3}\beta_i x_i\right)=0.183\ 2$$

$$\text{Prob}(y=3)=1-\Phi\left(\mu-\sum_{i=0}^{3}\beta_i x_i\right)=0.775\ 0$$

注意计算出来的3个类别的概率应该加总等于1。

与二分变量时一样,多项 Probit 和多项 Logit 的主要区别在于两者所依托的分布函数:多项 Probit 是正态分布,多项 Logit 是 Logistic 分布。两种分布之间的差别很小,除非数据规模很大或者处于分布尾端的观测数量非常重要。对模型估计结果的解释非常相似,因此,如何在两种方法上进行选择取决于实践的考量,例如,是否有现成的计算机程序或者是否有使用过某种方法的经验。

——Timothy M. Hagle

(李洋译　高勇校)

作者的注解:正如这里所讨论的,多项 Probit 涉及定序多项因变量。尽管如此,在一些资料中,通常使用"多项"这个词来指无顺序的或者分类因变量,并简单地用定序 Probit 来指定序数据。多项 Logit 是一种适合于对含分类因变量的模型进行估计的技术。

参考文献

Liao, T. F. (1994). *Interpreting probability models: Logit, probit, and other generalized linear models.* Thousand Oaks, CA: Sage.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics.* Cambridge, UK: Cambridge University Press.

McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, *4*, 103-120.

多个案研究(Multiple Case Study)

个案研究【Case Study】是一种着重理解单一场景的动态变化的研究方法。尽管这种方法可以用来进行描述和演绎(Yin, 1994),但是我们的关注点还是归纳性的理论发展,这种方法尤其适用于此。与汇总的统计研究相比,个案研究最大的优势在于对某种现象具体案例的深层次理解。

多个案研究是对同一现象进行两次或两次以上观测的个案研究的变式。它使得可重复性【Replication】成为可能——也就是说,使用多个个案来证实出现的构念【Constructs】和命题。它还能进行扩展研究,也就是说,使用多个个案来探索现象的不同侧面。结果就会形成更为稳健、一般化和成熟的理论。与多个案研究相反,单一案例可能包含更有启发意义的故事,但是产生高质量理论的可能性不大。

多个案方法的一个例子是葛路尼克和艾森哈特(Galunic & Eisenhardt, 2001)开展的一项归纳研究,他们对一家位列《财富》(*Fortune*)榜前100名的高技术企业中各部门如何获取10项经营特权进行了研究。通过研究不同部门获得新产品/市场份额的案例,作者识别并描述了这种现象的3种不同模式:新授权机遇、授权战争和名义授权。此外,如果不同案例具有共性且资料很丰富,那么研究者就能够在公司层次上发展出精妙的结构创新理论。

多个案研究从资料开始,以理论结束。由于演绎研究仅仅是对次序进行了颠倒,因此二者之间有相似性并不见怪。其相似性体现在都有一个事先定义的研究问题、明确选定的用于抽样的总体【Population(s)】、构念界定,以及在可能的时候用三角测量方法对构念进行测量。然而,尽管有许多相似点,但两者仍有些不同。

第一是抽样逻辑。与随机抽样【Random Sampling】不同,多个案研究使用的是理论抽样【Theoretical Sampling】。也就是说,我们选择个案的目的是满足概念分类、重复先前的研究发现或者拓展正在成形的理论(emergent theory)。此外,随着对哪些个案更能够验证和扩展正在成形的想法更加明确,在研究期间可以调整个案的选择。

第二是资料类型。与仅能使用定量数据不同,在这里可以使用定量和定性资料。实际上,由于那些来自各种媒介如观察、档案故事(archival stories)和访谈的定性资料仍然处于对案例进行深度理解的中心地位,因此,资料部分的重点一般来讲还是定性资料。由于重要的构念通常无法事先想到,这也是使用定性资料的另一个原因。此外,定性资料对于创造出一个潜在的、将建构与命题联系起来的理论逻辑是极为重要的。

第三是分析。与汇总观测值并进行统计检验不同,个案通常被看成所关注现象单独的事例,可以重复观测。我们的目标是实现数据和理论之间尽量紧密的匹配,因此分析就包括要不断从一个或几个个案中发展出试验性框架,然后在其他个案中进行检验(Glaser & Strauss, 1967)。这样的话,构念就提炼出来了,模式也出现了,同时构念之间的关系也得以确立。理论和资料之间的密切关系可以防止缺乏经验的学者常出现的研究者偏差问题。一旦粗略的理论模式出现,已有文献就可以不断纳入进来,进一步完善得到的理论。最终的目标是建构出高质量的理论。

——Filipe M. Santos
Kathleen M. Eisenhardt
(李洋译 高勇校)

参考文献

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*, 532-550.

Galunic, D. C., & Eisenhardt, K. M. (2001). Architectural innovation and modular corporate forms. *Academy of Management Journal*, *44*(6), 1229-1249.

Glaser, B., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Hawthorne, NY: Aldine de Gruyter.

Yin, R. K. (1994). *Case study research: Design and methods.* Thousand Oaks, CA: Sage.

多重分类分析(Multiple Classification Analysis, MCA)

多重分类分析(MCA)也称为析因方差分析(factorial ANOVA),是用来评估组间均值差异的一种定量【Quantitative】分析方法,组间均值差异还可根据相关因子和/或协变量及其效应来调整。MCA 的结果与使用虚拟变量【Dummy Variable】的多元回归【Multiple Regression】大致相同,尽管在信息的报告方式上有所差异。例如,在 SPSS 中,MCA 会生成一个方差分析【Analysis of Variance】表,包括适当的 F 检验【F Test】,模型解释的平方和【Sums of Squares】被分解为所研究的因素的贡献、协变量【Covariate】的贡献和设定的任何交互效应【Interaction】的贡献。F 检验要计算因子和协变量(如果进行了设定)所解释的平方和与未能解释的平方和之比,两者还要根据自由度进行调整,前者的自由度为因素类别 k 减 1 再加上协变量数目,后者的自由度为 N 减去前者的自由度。同时还会给出对每个因素和每个协变量的 F 检验。同时报告的还有一个表格,表中包括了各组根据因子和协变量调整后的预测均值,以及子组均值与总体均值的离差(也根据因子和协变量进行了调整)。在 SPSS 中运行因变量对一系列虚拟变量【Dummy Variable】和协变量的回归,也能生成含 F 检验的方差分析结果(但不会对可解释的平方和进行分解),此外,还有一组回归系数【Coefficient】,可以用这组回归系数来计算各组的均值,这些均值也已经根据模型设定的其他因素和协变量进行了调整。

表 1　多重分类分析的结果

		预测均值		离　差	
婚姻状况	n	未调整的	根据协变量调整	未调整的	根据协变量调整
已婚	3 243	8 814.432 6	8 737.775 8	903.789 1	827.132 3
丧偶	124	3 979.346 8	4 515.967 4	−3 931.296 7	−3 394.676 1
离异	210	4 696.409 5	4 818.359 6	−3 214.234 0	−3 092.283 9
独居	249	3 257.305 2	4 069.905 4	−4 653.338 3	−3 840.738 0
未婚	216	5 087.324 1	4 878.871 6	−2 823.319 4	−3 035.771 9

举例说明,考虑婚姻状况和受教育年限(协变量)对家庭总收入的影响。表 1 显示了用 SPSS 进行统计的 MCA 输出表格,这个统计使用了全国纵贯调查(National Longitudinal Surveys)中 1967 年年龄在 30~45 岁的成年女性数据。家庭年收入的总体均值是 7 910.64 美元。第一列显示了婚姻状况类别,第二列显示了每组的个案数,第三列和第四列显示了未调整和调整后的预期均值。未调整均值就是各组的实际均值(即在每个婚姻状况类别中个案的平均家庭收入)。

每个婚姻状况类别的调整后预期均值就是在控制了受教育年限之后的预期均值。表 1 的离差列显示了子组均值和全体被调查者总体均值的比较。其中,未调整离差表明已婚人士的平均家庭收入比总体均值 7 911美元高出约 903 美元。这一差异中有 76 美元是由于已婚人士的平均受教育年数更高,而受教育年数与家庭收入存在正相关关系引起的。

如果我们对回归方程进行估计,方程中家庭总收入是因变量,受教育年限和一组表示不同婚姻状况差异的虚拟变量是自变量,我们就可以得到同样的对模型拟合的 F-检验。此外,我们只要把平均受教育年限和适当的虚拟变量值代入方程,就可以使用回归系数来再现调整后的组均值。

——Melissa A. Hardy

Chardie L. Baird

(李洋译 高勇校)

参考文献

Andrews, F. M., Morgan, J. N., & Sonquist, J. A. (1967). *Multiple classification analysis: A report on a computer program for multiple regression using categorical predictors*. Ann Arbor, MI: Survey Research Center, Institute for Social Research.

Retherford, R. D., & Choe, M. K. (1993). *Statistical models for causal analysis*. New York: John Wiley.

多重比较(Multiple Comparisons)

多重比较比较的是某一分类变量的 J 个均值下的样本均值($J \geqslant 3$)。单因素方差分析【One-Way ANOVA】、固定效应模型【Fixed-Effects Model】只是判断均值之间是否存在显著差异,但并不判定哪些均值之间的差异是显著的。

J 个均值之间的比较可以用以下均值线性组合 ψ 来表示:

$$\psi = c_1\mu_1 + c_2\mu_2 + \cdots + c_J\mu_J = \sum_{j=1}^{J} c_j\mu_j \quad (1)$$

其中 c 的加总等于零,但是并非每个 c 都是零。为了对 ψ 进行估计,可以使用

$$\hat{\psi} = c_1\bar{Y}_1 + c_2\bar{Y}_2 + \cdots + c_J\bar{Y}_J = \sum_{j=1}^{J} c_j\bar{Y}_j \quad (2)$$

其中,c 与 ψ 是给定的 c 相同。这里可以有许多不同的比较,因此被称为多重比较。对多重比较进行检验(Multiple Comparison Procedures, MCP)的历史可以追溯到罗纳德·费希尔的早期工作中,他提出在 ANOVA 的 F 检验显著之后需要进行单独的 T 检验:这一想法在 Fisher-Hayter MCP 的方法中得以体现并进一步完善,这在接下来的论述中将有所涉及。亨利·谢菲和约翰·图基以及其他人进一步发展了多重比较技术,形成了他们各自的多重比较方法,也使得多重比较成为应用文献中的重要主题。

下面会先简要介绍多重比较中常用的基本概念，然后会给出几种不同的多重比较程序，最后介绍最新的发展情况。

基本概念

对于 J 个均值中的两个均值进行成对比较(a pairwise comparison)时，这两个均值的权重是 1 和-1，其他均值的权重是 0，这就成了两个均值之差。例如，在 $J=4$ 时，均值 1 和均值 2 的成对比较可以表示为

$$\hat{\psi}_1 = \bar{Y}_1 - \bar{Y}_2 \quad (3)$$

其中，$c_1 = 1, c_2 = -1$，其他的 c_s 等于零。非成对比较就是其他非成对比较的比较，例如

$$\hat{\psi}_2 = \bar{Y}_1 - \frac{\bar{Y}_2 + \bar{Y}_3}{2} \quad (4)$$

其中，$c_1 = 1,\ c_2 = c_3 = -1/2$，且 $c_4 = 0$。

计划(或事先)比较(planned comparisons)是研究者在获取结果之前就计划进行的比较。计划比较通常是基于指导研究项目的理论预测。事后比较(post hoc comparisons)是在获取结果之后再进行计算。在事后比较中，研究人员可以基于结果来选择比较。事后比较的研究结论可能有也可能没有理论基础。

对于任何两个比较来说，都要考虑多重比较的正交性。在所有样本的样本量都是 n 时，如果满足以下条件，那么比较 1 和比较 2 是正交的：

$$\sum_{j=1}^{J} c_{1j}c_{2j} = 0 \quad (5)$$

对于 J 个均值来说，最多只有 $J-1$ 对比较是彼此正交的。

如果 ANOVA 假定【ANOVA Assumptions】满足，那么样本比较的抽样分布【Sampling Distribution】的均值、方差、方差估计值、形态(正态)就都可以知道。给定上述信息，可以对任何比较构建一个 t 统计量：

$$t_{\hat{\psi}} = \frac{\hat{\psi} - \psi}{\sqrt{MS_W \sum_{j=1}^{J} \frac{c_j^2}{n_j}}} \quad (6)$$

如果是成对比较且 n 相等，t 统计量可以简化为

$$t = \frac{\bar{Y}_j - \bar{Y}_k}{\sqrt{\frac{MS_W(2)}{n}}} \quad (7)$$

其自由度 $df_W = J(n-1) = N-J$。

如果仅对其中的一个比较感兴趣，并且也满足通常的 ANOVA 假设，就可以用 t 统计量来进行假设检验：

$$H_0: \psi = 0 \quad (8)$$

这里要引入 $df=df_W$ 的 t 分布的一个临界值。对于多重比较来说，会有类似于假设(8)的多重假设，会用到与所进行的多重比较相适宜的 T 统计量的临界值。这样的检验通常大多是双尾检验【Two-Tailed Tests】。

下面介绍的所有多重比较程序都能让研究者进行假设检验。但是，如果研究者希望得到 ψ 的区间估计，有些 MCP 就不适用了，因为它们无法给出总体比较的区间估计。对于那些能够计算出置信区间【Confidence Interval】的方法来说，这个区间是

$$\hat{\psi} - \text{crit}\sqrt{\frac{MS_W}{n}\sum_{j=1}^{J} c_j^2} \leqslant \psi \leqslant \hat{\psi} +$$

$$\text{crit}\sqrt{\frac{MS_W}{n}\sum_{j=1}^{J}c_j^2} \quad (9)$$

其中,crit 如同在 t 分布中那样的 α 水平下的一个双尾临界值。对于那些既需要假设检验又需要进行区间估计的研究者来说,可以使用置信区间来检验比较结果为零的假设,如果区间不含零则拒绝 H_0。

研究者首要关注的是 α-控制和权重[参见 Alpha(α)【Alpha】、检验效力【Power of The Test】和第一类误差【Type I Error】]。这里有两类基本的 α-控制(误差率):每一次比较的 α-控制和某一组比较的 α-控制。选择何种类别的 α-控制,决定了 α'在每一次比较上的值,并影响最终统计量的效力。不同多重比较程序的相对效力可以通过比较临界值来评估,要选择临界值最小的多重比较程序。

为每次比较设定 $\alpha'=\alpha$(通常是0.05),然后选择一个 α 水平的 t 临界值,这被称为单次比较误差率(Error Rate Per Comparison, ERPC)。用 ERPC 来控制 α 的问题在于,至少有一次比较中出现第一类误差的概率 p 就会随着比较(C)的数目增加而增加。对于 α' 水平下的 C 个正交比较来说,可以证明至少有一次比较中出现第一类误差的概率为 $p=1-(1-\alpha')^C \leqslant C\alpha'$。如果 C 个比较并非正交,那么,至少有一次比较中出现第一类误差的概率为 $p \leqslant 1-(1-\alpha')^C \leqslant C\alpha'$,但是经研究表明 p 会非常接近于 $1-(1-\alpha')^C$。

ERPC 中至少有一次比较中出现第一类误差的概率 p 较大,但是效力很强。与其他任何误差率控制方法相比,采用 ERPC 控制误差率的效力更高,但代价是至少有一次比较中出现第一类误差的概率 p 也更高。基本假设检验中有 α 和效力之间的平衡问题,与此类似,允许 p 值更高则效力也更高。

α 控制的第二个基本类别是对一组比较控制其 α。对一组比较控制其 α 的方法又有两种,其中一种是单组比较误差率【Error Rate Per Family】。与其他误差率定义不同,ERPF 并不是一个概率,它是一组比较中陈述错误(即错误拒绝原假设的第一类误差)的平均数。通过将 α'设为一个小于 α 的值可以得到 ERPF,最简单的方法是设 α'加起来等于 α。例如,如果每次比较的 α'值都相等,那么对于 α'水平下的 C 个正交比较来说,ERPF 等于 $C\alpha'$。如下所示,这种控制误差率的方法体现在 Dunn 方法的逻辑中。

第二种对一组比较进行 α-控制的方法是按组比较误差率(Error Rate Family Wise, ERFW)。ERFW 是一个概率,定义为至少有一次比较中出现第一类误差的概率 $p=1-(1-\alpha')^C$。ERFW 和 ERPF 在定义和概念上不同,但是在实践中多重比较如果满足了其中一种,通常就能满足另一种。因此,研究者关于 α-控制的基本决策,实际上是决定要用对每一次比较的误差率进行 α-控制(ERPC),还是对某一组比较进行 α-控制(ERPF 或 ERFW)。

为了说明 3 种不同类型 α-控制之间的关系,设每个 C 比较都引入 α':采用 ERPC,则 $\alpha=\alpha'$,采用 ERFW,则 $\alpha<1-(1-\alpha')^C$,采用 ERPF,则 $\alpha<C\alpha'$。如果将上述关系叠加,则有 $\alpha' \leqslant 1-(1-\alpha')^C \leqslant C\alpha'$;因此,ERPC<ERFW<ERPF。对于 $C=5$ 的正交比较来说,如果设 $\alpha'=0.01$,那么 ERPC=0.01,ERFW=0.049,ERPF=0.05。

各类假设考虑的总体均值中存在的样式,主要关注的是均值是否相等。如果所有 J 组的均值都相等,那么总体虚无假设就成立。这一虚无假设要通过总体 F 检验进行验证,因此,被称为总体虚无假设(overall null hypothesis)或者"完全(full)"虚无假设。如果总体虚无假设不存在,但是存在 1 个以上均值相等的子集,那么多重虚无假设

就成立。子集内的均值必须相等,但是子集间的均值必须不同。这样每个子集就表示了一个虚无假设。如果有 M 个均值相等的子集,就有 M 个多重虚无假设。例如,如果 $J=4$,均值 1、均值 2 相等且与均值 3、均值 4 不同,但均值 3 和均值 4 相等,那么就有 $M=2$个虚无假设。如果总体虚无假设不存在但一些组的均值相等,那么部分虚无假设成立。比如,在 $J=4$ 的情况下,如果前 3 个均值相等,但是第 4 个均值比前 3 个均值更大,那么在 3 个相等的均值中部分虚无假设成立。

逐步多重比较(stepwise MCP)中,一个检验部分地依赖于另一个检验。在逐步 MCP 中,只有总体 F 显著时才可以对各种比较进行计算,或者只有其他比较显著时才可以对某个比较进行检验。没有形成上述依赖关系的多重比较称为非逐步检验方法或同步检验方法(Simultaneous Test Procedures, STP),因为这些检验可以同步进行。

多重比较检验

许多 MCP 出现在单因素 ANOVA 中。如果你使用 T 统计量,并已经决定了如何控制误差率,那么,选择多重比较检验就只需要选择临界值(理论分布)。除非另作说明,置信区间可以根据每个多重比较检验的临界值计算得出。

选择 ERPC 控制 α,就是常见的多重比较 t 检验,如果下式成立就拒绝 H_0,

$$|t_{\hat{\psi}}| \geq t^{\alpha}_{df_W} \tag{10}$$

否则就要保留 H_0。注意常见的 t 统计量既可以用于成对比较,也可以用于非成对比较。

如果计划进行 C 个比较,并且选择 ERPF 或者 ERFW 控制 α,那就用 α 除以 C 得到 $\alpha'=\alpha/C$,用于单次比较,根据 Dunn 多重比较检验(参见 Bonferroni 技术【Bonferroni Technique】),如果下式成立则拒绝 H_0,

$$|t_{\hat{\psi}}| \geq t^{\alpha'}_{dfW} \tag{11}$$

否则就保留 H_0,Dunn 多重比较检验可用于成对或非成对比较。

Tukey 多重比较检验通过引入来自 J 个均值、组内自由度为 df_W 的学生化极差分布(studentized range distribution)中 α 水平的临界值来控制 α,因此,α-控制是通过学生化极差分布在所选的按组比较误差率 α 水平下的临界值来实现的。根据 Tukey 多重比较检验,如果下式成立则拒绝 H_0,

$$|t_{\hat{\psi}}| \geq \frac{q^{\alpha}_{J,df_W}}{\sqrt{2}} \tag{12}$$

否则就保留 H_0。Tukey 多重比较检验可以用于成对比较。

Scheffé 多重比较检验通过引入来自组间和组内自由度为 df_B和 df_W的 F 分布中 α 水平的临界值来实现 α 控制,所以,α 控制是通过 F 分布在所选的按组比较误差率 α 水平下的临界值来实现的。根据 Scheffé 多重比较检验,如果下式成立则拒绝 H_0,

$$|t_{\hat{\psi}}| > \sqrt{(J-1)F^{\alpha}_{J-1,df_W}} \tag{13}$$

否则就保留 H_0,在采用 ERFW 来进行 α 控制时,Scheffé 方法差不多可以用于任意类型的比较:成对比较、非成对比较、正交多项式等。

只有当总体方差分析的 F 检验显著后,才可用 Fisher-Hayter 多重比较检验进行检验。此时,选择 ERFW 控制 α,临界值为 q,$J-1$是均值数量参数。判定的方法是进行总体方差分析的 F 检验,如果在 α 水平下 F 检验显著且下式成立,则拒绝 H_0,

$$|t_{\hat{\psi}}| \geq \frac{q^{\alpha}_{J-1,df_W}}{\sqrt{2}} \quad (14)$$

否则就保留 H_0。Fisher-Hayter 多重比较检验是一种逐步成对比较法；由于其临界值并未考虑第一步检验总体 F 显著的情况，因此它不能进行置信区间的计算。

Ryan 多重比较检验综合了多个统计学家的相关研究而得出。首先，将 J 个均值按顺序排列。这里要用到比较的延展度(stretch size)这个概念，它是在比较的两个均值之间且包括这两个均值的有序均值(ordered means)数目。延展度用 p 表示。Ryan 多重比较检验采用 ERFW 进行 α 控制，临界值为 q，p 是均值数量参数，其中 α 受 p 的影响而取值不同。对于 Ryan 多重比较检验来说，如果比较中的均值并未包含在之前保留的假设的延展内且下式成立，那么就拒绝 H_0，

$$|t_{\hat{\psi}}| \geq \frac{q^{\alpha_p}_{p,df_W}}{\sqrt{2}}$$

其中

$$\alpha_p = \alpha \quad 对于\ p = J, J-1$$
$$\alpha_p = 1 - (1-\alpha)^{\frac{p}{J}} \quad 对于\ p \leq J-2 \quad (15)$$

否则，就保留 H_0。Ryan 多重比较检验是逐步成对比较法；由于其临界值并未考虑之前步骤中的检验显著性情况，因此，它不能进行置信区间计算。

Games-Howell(简称 GH)多重比较检验可以同时兼顾 n 不相等和方差不相等的情况。这种检验方法用 ERFW 能够很好地控制 α，它对于每个 $j \neq k$ 的均值来计算下列统计量：

$$t_{jk} = \frac{\bar{Y}_j - \bar{Y}_k}{\sqrt{\frac{s_j^2}{n_j} + \frac{s_k^2}{n_k}}} \quad (16)$$

如果下式成立，则拒绝 H_0，

$$|t_{jk}| \geq \frac{q^{\alpha}_{J,df_{jk}}}{\sqrt{2}} \quad (17)$$

否则就保留 H_0，其中

$$df_{jk} = \frac{\left(\frac{s_j^2}{n_j} + \frac{s_k^2}{n_k}\right)^2}{\frac{\left(\frac{s_j^2}{n_j}\right)^2}{n_j - 1} + \frac{\left(\frac{s_k^2}{n_k}\right)^2}{n_{k-1}}} \quad (18)$$

在实际操作中，建议将 df_{jk} 四舍五入为其最接近的整数(the whole number)，作为临界值 q 的自由度。GH 多重比较检验可用于成对比较。要了解更多基本的多重比较检验，可以参考 Toothaker(1993)。关于多重比较检验的数学解释可以参考 Hochberg & Tamhane(1987)，Miller(1981)。

最新进展

追随图基的研究，多位研究者研究了样本均值、样本方差和均值检验这些统计量对于极端值的敏感性，研究它们是否缺乏稳健性。这些研究重点考察了总体服从重尾分布时的稳健性缺失问题。重尾分布偏离了正态性，因此，往往会导致均值(以及方差)的方差堆积、α 控制缺失、均值检验效力较低等问题。近年来，多重比较检验领域的相关研究，主要是通过引入 20% 截尾后均值(trimmed means)和 20% 缩尾后方差(winsorized variances)来生成 t 统计量。首先，计算各尾在 $g = [0.2n]$ 时的截取值，其中

[0.2n]表示取 0.2n 的整数部分。对于 J 组的每个均值来说,计算 20%截尾均值 $\bar{Y}_t$ 的步骤包括:从样本各尾中分别修整【Triming】掉前 g 个最大的极值,将剩余的 $h=n-2g$ 个值加总并除以 h。这些 20%截取均值的方差,要通过缩尾方差函数进行估计。而要计算缩尾方差,需要再看 J 个样本的各尾上的 g 个极值。设 $Y_{(1)} \leqslant Y_{(2)} \leqslant \cdots \leqslant Y_{(n)}$,即每组均值按取值排序。通过下式能计算出缩尾均值:

$$\bar{Y}_w = \frac{(g+1)Y_{(g+1)} + Y_{(g+2)} + \cdots + Y_{(n-g-1)} + (g+1)_{(n-g)}}{n} \tag{19}$$

同时缩尾方差也能通过下列公式计算得到:

$$S_w^2 = [(g+1)(Y_{(g+1)} - \bar{Y}_w)^2 + (Y_{(g+2)} - \bar{Y}_w)^2 + \cdots + (Y_{(n-g-1)} - \bar{Y}_w)^2 + (g+1)(Y_{(n-g)} - \bar{Y}_w)^2]/(n-1) \tag{20}$$

这样通过下面的公式,可以得到每组的截尾均值的方差估计值:

$$d = \frac{(n-1)S_w^2}{h(h-1)} \tag{21}$$

稳健成对检验统计量是

$$t_w = \frac{\bar{Y}_{tj} - \bar{Y}_{tk}}{\sqrt{d_j + d_k}} \tag{22}$$

其中估计的组内自由度是

$$df_w = \frac{(d_j + d_k)^2}{\dfrac{d_j^2}{h_j - 1} + \dfrac{d_k^2}{h_k - 1}} \tag{23}$$

在不同类型的多重比较检验中,例如,上面提到的 Ryan 多重比较检验等,都可以使用自由度为 df_W 的稳健统计量 t_W。更多关于稳健估计的资料可参考 Wilcox(2001)。关于稳健估计以及相应的多重比较检验方法,可参考 Keselman, Lix, & Kowalchuk (1998)。

——Larry E. Toothaker

(李洋译 高勇校)

参考文献

Hochberg, Y., & Tamhane, A. C. (1987). *Multiple comparison procedures*. New York: John Wiley.

Keselman, H. J., Lix, L. M., & Kowalchuk, R. K. (1998). Multiple comparison procedures for trimmed means. *Psychological Methods*, *3*, 123-141.

Miller, R. G. (1981). *Simultaneous statistical inference* (2nd ed.). New York: Springer-Verlag.

Toothaker, L. E. (1993). *Multiple comparison procedures*. Newbury Park, CA: Sage.

Wilcox, R. (2001). *Fundamentals of modern statistical methods: Substantially improving power and accuracy*. New York: Springer.

多重相关(Multiple Correlation)

多重相关来源于多元回归分析【Multiple Regression Analysis】;它用一个数

字来概括回归模型中得到预测值的准确性。

多元回归模型用一组自变量(预测变量)来预测某一个因变量 Y(或效标)。多元线性回归【Linear Regression】(或最小二乘【Ordinary Least Squares】回归分析)是多元回归最常见的形式,其中自变量根据如下线性回归方程组成了一个线性组合:

$$\hat{Y} = b_1X_1 + b_2X_2 + \cdots + b_pX_p + b_0$$

其中,X 是预测变量。每个预测项乘以一个权重,此权重称为偏回归系数【Partial Regression Coefficient】$b_1, b_2, \cdots, b_p$。根据回归方程,预测变量得分的线性组合(或加权加总)可以计算得出。加权加总为 $\hat{Y}$,这称为预测值。回归系数 $b_1, b_2, \cdots, b_p$ 要使得实际因变量 Y 和预测值 $\hat{Y}$ 之间的相关性尽可能强。效标 Y 和一组 X 的线性组合之间的最大相关系数就是多重相关系数 $R_{Y\hat{Y}}$。在日常研究实践中,通常提供的是相关系数的平方,它被称为多重相关系数平方 $R^2_{Y\hat{Y}}$,也被称为 R 方【R-Squared】。

R 方是多元回归分析的核心测度,它概括了预测变量组在解释因变量时的整体准确度(overall adequacy)。R 方是预测变量组所解释的因变量的变差比例。

例如,一门本科生统计学课程在一学期中进行了 3 次测试。假设一个班级有 50 名学生,我们使用线性回归的分析方法,根据测试 1 和测试 2 的分数来预测学生在测试 3 中的表现,所得到的线性方程如下所示:

$$\hat{Y} = 0.15 \times 测试1 + 0.54 \times 测试2 + 28.91$$

对每名学生来说,我们将测试 1 和测试 2 的得分代入回归方程,计算出测试 3 中的预测值 $\hat{Y}$。然后,计算预测得分和测试 3 的实际得分之间的相关系数,得到的就是多重相关系数。这里多重相关系数 $R_{Y\hat{Y}} = 0.65$,说明相关性较强,学生在测试 3 中的表现与第一次和第二次测试密切相关。多重相关平方 $R^2_{Y\hat{Y}} = 0.42$。我们可以这样对结果进行描述,测试 3 的得分中,有 42% 的变差可以由测试 1 和测试 2 的得分来进行解释。

多重相关系数的取值范围为 0~1。随着预测变量逐渐加入回归方程中,多重相关或者保持不变,或者增强。多重相关系数并不考虑预测变量的数量。样本中计算出的多重相关系数一般都会高估总体的多重相关系数。调整后的 R 方则偏差较少(尽管也有偏差),其公式如下,n 是样本数量,p 是预测变量数:

调整后的

$$R^2_{Y\hat{Y}} = 1 - (1 - R^2_{Y\hat{Y}}) \times \frac{n-1}{n-p-1}$$

在我们的分析中,$n = 50$ 名学生,$p = 2$ 个预测变量,调整后的 $R^2_{Y\hat{Y}} = 0.39$。

这里提到的多重相关系数只适用于线性回归分析。在其他形式的回归分析中(如 Logistic 回归【Logistic Regression】或者泊松回归【Poisson Regression】),也有一些与多重相关系数(以及多重相关系数的平方)可类比的测度;但严格地讲,它们并不是这里所定义的多重相关系数。

——Leona S. Aiken

(李洋译　高勇校)

*也可参考多元回归分析【Multiple Regression Analysis】、回归分析【Regression Analysis】和 R 方【R-Squared】。

参考文献

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Draper, N. R., & Smith, H. (1998). *Applied*

regression analysis (3rd ed.). New York: John Wiley.

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear regression models* (3rd ed.). Chicago: Irwin.

多重对应分析(Multiple Correspondence Analysis)

参见对应分析【Correspondence Analysis】。

多元回归分析(Multiple Regresion Analysis)

多元回归分析包括一系列将一组自变量【Independent Variables】(预测变量)与一个因变量【Dependent Variable】(效标)关联起来的统计方法。多元回归分析(MR)的结果之一是多元回归方程,它在考虑方程中所有其他自变量的前提下测量每个预测变量与效标之间的关系。多元回归分析应用很广泛。它可以进行描述分析,对一组预测变量与效标的关系进行简单总结。它也可以用来预测【Prediction】,还可以用来建构模型【Model】和验证理论。

多元回归的结构

多元回归最常见的形式为多元线性回归【Linear Regression】分析,其多元回归方程如下:

$$\hat{Y}= b_1X_1+ b_2X_2+\ldots+b_pX_p+ b_0$$

X 是预测变量。每个预测变量都乘以一个权重,即偏回归系数【Partial Regression Coefficien】,b_1, b_2, $\cdots$, b_p。方程的截距 b_0被称为回归常数或者回归截距【Intercept】。在回归方程中,预测变量组的取值合并为某一个预测值 $\hat{Y}$,它是所有预测变量的线性组合。

自变量和因变量分别被称为预测变量和效标,顾名思义,多元回归被用来根据不同个体的自变量的取值来预测个体在效标上的取值。例如,高校招生办公室可以建立一个多元回归方程,通过下面 3 个预测变量来对高校新生平均绩点(效标)进行预测:高中年级平均成绩以及高校入学考试中的语文和数学成绩。基于申请入学者的这些高中成绩,就可以预测出这些申请人可能取得的新生平均绩点。

在多元回归中,回归方程中每个预测变量的偏回归系数【Partial Regression Coefficient】实际上考虑了这个预测变量与效标的关系、除上述预测变量外其他所有预测变量与效标的关系、所有预测变量之间的关系。偏回归系数给出了这一预测变量对整体预测的单独贡献,这种贡献排除了方案中包括的所有其他解释变量,也就是说假定所有其他解释变量不变。

多元回归给出一个汇总统计量来表明预测变量组能多么精准地对效标进行再现(reproduce)或者"解释",这个数字被称为多元相关【Multiple Correlation】平方,表示为 R^2,$R^2_{Y\hat{Y}}$或 $R^2_{Y.1,2,\cdots,p}$。其中 $R^2_{Y\hat{Y}}$表示多元相关平方是预测值和效标值的相关系数的平方。事实上,多元相关是预测变量的线性组合和效标之间可能取得的最大相关值。第三种表示方式 $R^2_{Y.1,2,\cdots,p}$说明这种相关关系反映了预测变量组 $1,2,\cdots,p$ 与效标之间的总体关系。

多元相关平方测量了预测变量组所能解释的效标变异(variation in the criterion)比例,是对全体回归分析的效应量【Effect Size】的测量。

回归系数值是依据将预测误差最小化来确定的。对每个个体来讲,我们都能测量出预测误差,即个体在效标上的观测值和预测值之间的差异 $(Y-\hat{Y})$,也称作残差。在一般最小二乘【Ordinary Least Squares】回归分析(最常见的估计【Estimation】回归系数的方法)中,选择偏回归系数要以残差平方和最小为目标,即一般最小二乘(Ordinary Least Squares,OLS)准则。得到的 OLS 回归系数,同时也能够使得来自预测变量组的多元相关最大化。

建构回归模型

这里有一个假定的健康心理学案例,可以用来说明回归分析的各个方面以及它在模型建构中的作用。我们想要建立一个由皮肤癌预防有关的因子组成的模型。目标总体是居住在西南部或东北部地区的年轻的成年白种人。作为效标,对他们保护自己以免罹患皮肤癌的意图进行了测量,建立起10分的测量尺度(称为"意图")。这里有4个预测变量:(a)由于肤色等客观原因而罹患皮肤癌的客观风险(称为"风险");(b)参与者居住的区域是西南部还是东北部(称为"区域");(c)参与者对晒黑的好处的评价(称为"好处");(d)参与者对关于防晒的指令性规范(即他或她"应该"如何防晒)的认知(称为"规范")。假设风险是防晒意图的正向预测变量,而晒黑的好处是一种负向的预测变量。期望由于西南部地区日晒程度较强因而有较高的防晒意图。此外,也不确定指令性规范的影响,毕竟年轻人可能并不会遵循保护其健康的各种建议。

通过选择预测变量的过程可以看到多元回归的灵活性。"风险""好处"和"规范"都是连续变量【Continuous Variable】,按6分尺度进行测量。区域是一个分类的【Categorical】变量。分类预测变量需要特殊的编码方式,以此表明它们是分类的而非连续的。在实践中,通常使用虚拟编码(虚拟变量编码【Dummy Variable Coding】)、效应编码【Effect Coding】和对比编码【Contrast Coding】3种编码方式。对于二分(两个类别)变量的虚拟编码来说,通常在一个预测变量上将一个类别设定为1,其他类别设定为0。这里采用虚拟编码,设西南部地区=1,东北部地区=0。

表1a显示了预测变量和效标的相关【Correlation】关系,$n=100$。相关矩阵的最后一列显示了每个预测变量与效标的相关性。得到的关系是符合我们期望的。同时,还注意到几个预测变量之间存在着显著【Significant】相关。"风险"与"好处"呈负相关。标准与区域呈正相关。由于西南部地区编码为1,东北部地区编码为0,因此,正相关的结果表明,在西南部地区,应注意防晒的指令性规范更深入人心。

表1b是一系列的二元回归方程(即一个预测变量和一个效标)。每个预测变量有一个方程,以及多元相关平方 R^2。每个预测变量方程中的回归系数测量的是每个预测变量改变一个单位时 Y 的变化。例如,"风险"在10分尺度上每增加1分,"意图"增加0.145分,对于转换为虚拟变量的区域来说,回归系数显示,西南部地区(编码为1)"意图"变量的算术平均值要高于东北部地区1.049分。在二元回归中,多元相关平方就是指预测变量和效标相关系数的平方值。我们注意到,"风险"对于"意图"有统计显著性。"风险"单独解释了 Y 变量3.7%(也就是 $R^2=0.037$)的变异。同样,我们注意到"好处"变量也具有统计显著性,它解释了 Y 变量14.0%的变异。

表 1　多元回归分析图

a. 所有预测变量与效标之间的相关关系

	风　险	区　域	好　处	规　范	意　图
风　险	1.000	− 0.075	− 0.203*	− 0.094	0.193*
区　域	− 0.075	1.000	0.168 +	0.424*	0.313*
好　处	− 0.203*	0.168 +	1.000	0.254*	−.374*
规　范	− 0.094	0.424*	0.254*	1.000	0.028
意　图	0.193*	0.313*	− 0.374	0.028	1.000

b. 二元回归分析:单一预测变量回归方程

1.仅风险	$\hat{Y} = 0.145 \times$ 风险* + 3.991	$R^2 = 0.037$
2.仅区域	$\hat{Y} = 1.049 \times$ 区域* + 4.407	$R^2 = 0.098$
3.仅好处	$\hat{Y} = -0.474 \times$ 好处* + 6.673	$R^2 = 0.140$
4.仅规范	$\hat{Y} = 0.044 \times$ 规范 + 4.760	$R^2 = 0.001$

c. 根据风险和好处的预测

$\hat{Y} = 0.180 \times$ 风险 $- 0.443 \times$ 好处* + 5.985　　$R^2 = 0.154$

d. 回归分析:多元相关平方检验,风险和好处

方差来源	平方和	自由度	均方	F 比	p
回归	43.208	2	21.606	8.858	0.001
残差	236.582	97	2.439		

e. 完全模型:所有 4 个预测变量的预测

$\hat{Y} = 0.202 \times$ 风险 $- 0.516 \times$ 好处* + 1.349 × 区域* − 0.039 × 规范 + 5.680　　$R^2 = 0.304$

f. 截尾模型和抑制示例:关于风险和区域的预测

未标准化方程	$\hat{Y} = 0.319 \times$ 风险* + 1.104 × 区域* + 3.368	$R^2 = 0.145$
标准化方程	$\hat{Y} = 0.218 \times$ 风险* + 0.329 × 区域*	$R^2 = 0.145$

注:* $p < 0.05$。

现在我们采用"风险"和"好处"作为一个多元回归方程中的两个预测变量,如表 1c 所示。以下两个方面要引起注意。首先,与二元回归方程不同的是,"风险"不再是"意图"的显著性预测变量。这是因为(a)与"风险"相比,"好处"与"意图"相关性更强,(b)上述前两个预测变量之间同样存在相关关系($r = -0.203$)。其次,所有预测变量共同解释了 15.4%的方差(即 $R^2 = 0.154$),这要低于每个预测变量单独对变异

所能解释的比例的总和,即 3.7%+14.0% = 17.7%。这是由于"风险"和"好处"之间在解释意图时产生了冗余(redundancy)。

在多元回归中,将效标 Y 的总体变异(通常记作 SS_Y)分为互不重叠的(也称为正交的)两个部分。第一个是可预测的变异或者是回归平方和【Regression Sums of Squares】(通常记作 $SS_{回归}$),它对回归方程在多大程度上再现了不同个体在效标上的差异进行了总结。准确地说,它是预测值偏离均值的变异。第二个是残差变异或残差平方和【Residual Sum of Squares】(记作 $SS_{残差}$),它是效标变异 SS_Y 中无法被预测变量组解释的那一部分。多元相关平方可以简单地用 $SS_{回归}/SS_Y$ 的比值表示,或者说是变异中能被预测变量组所解释的比例。这些变异的来源可以参见表 1d 这个回归汇总分析表。变异的两个来源可以用来检验整体性的虚无假设【Null Hypothesis】,即总体中多元相关平方是否为零的检验。在本例中,表中的 F 比【F Ratio】表明可以拒绝虚无假设,即用"风险"和"好处"这两个预测变量可以解释"意图"的很大一部分变异。

我们是否可以通过增加其他的预测变量来增强预测效果呢?层次回归分析(hierarchical regression analysis)回答了这个问题。层次回归分析需要估计一系列的回归方程,这些回归方程的预测变量组(每组一个或多个变量)是先后加入回归方程中的,每加入一个预测变量组时都要检验预测的准确性是否得到了提升(即多元相关平方是否增加了)。如表 1e 所示,在回归方程中增加了"区域"和"规范"变量,生成了含 4 个预测变量的回归方程,最终的模型包括所有的预测变量。可以看到,在由 4 个预测变量组成的回归方程中,"好处"和"区域"对于意图变量来说都是显著性变量。此外,"规范"变量的回归系数接近于零并不让我们感到奇怪,因为它与"意图"并不相关。我们注意到多元相关平方有了显著增长,从仅考虑"风险"和"好处"两个预测变量时的 $R^2=0.154$,增长到考虑所有 4 个预测变量时的 $R^2 = 0.304$。增加的 0.150×(0.304 - 0.154=0.150)是排除了"风险"和"好处"两个变量之后(即假定"风险"和"好处"不变时)"区域"和"规范"两个变量所预测的效标的变异比例。在这个例子中,R^2 的增加是统计显著的。包含 4 个预测变量的最终回归方程就是对预防皮肤癌的意图的模型。在解释时要多加留意的是,由于"风险"与"意图"没有相关关系,因此,"风险"并不是"意图"的统计显著的预测变量。此外,还必须解释"风险"与"好处"的关系,后者是"意图"的另一个重要的预测变量。

在多元回归分析中,预测变量之间的相互作用可能会导致意想不到的结果。以表 1f 为例,通过"风险"和"区域"两个变量对"意图"变量进行预测。从二元回归中可以看到,仅考虑风险的话,$R^2 = 0.037$;仅考虑区域的话,$R^2=0.098$。但是,如果同时考虑风险和区域,$R^2=0.145$,比单独考察两个预测变量的值要大。这是一个特殊的(不常见的)结果,被称作多元回归中的抑制。通常认为,之所以形成抑制效应【Suppressor Effect】,是因为一个预测变量将另一个预测变量的部分不相关方差剔除出去,因此,改善了另一个预测变量的预测效果。

到目前为止,考察的回归系数都是非标准化的回归系数。这样的回归系数取决于不同预测变量特定的测量单位,而第二组回归系数,即标准化回归系数【Standardized Regression Coefficients】则是建立在所有预测变量和效标的标准化值(均值为 0、标准差为 1 的 z 值)的基础上。标准化回归系数

的测量单位相同(即标准差单位)。表 1f 显示了关于“风险”和“区域”的标准回归方程。“风险”变量的系数是 0.218,意味着“风险”变量的标准差增长 1 个单位,“意图”的标准差增长 0.218 个单位。

迄今为止,仅仅考察了每个预测变量与效标的线性关系。在多元线性回归中,同样可以对非线性【Nonlinear】关系进行处理,可采用下面的任意一种方法:多项式【Polynomial】回归以及预测变量和/或效标转换【Transformations】。多项式回归是在线性回归方程的基础上生成一个包括预测变量较高阶函数的回归方程,方程的形式可以是二次方程(一次弯曲)关系,即 $\hat{Y} = b_1X + b_2X^2 + b_0$,或者是三次方程(二次弯曲)关系,即 $\hat{Y} = b_1X + b_2X^2 + b_3X^3 + b_0$,等等。假设我们认为防晒的意图会随着习惯性暴露于日晒的水平提升而增强,但只是增强到一定程度。如果习惯性暴露于日晒的水平非常高,那么个体也就会忽略防晒,因此,在高水平日晒中,防晒的意图也会下降。为了验证这个“倒 U 形”关系的假设(意图先是上升,然后随着习惯性日晒水平的提升而下降),对一个二次方程【Quadratic Equation】的 X^2 项的系数 b_2 进行了考察。对于“倒 U 形”关系来说,这个系数应该为负。假设 b_2 为零,这就表明了二次(“倒 U 形”)关系并不成立。对于一个“U 形”关系来说,b_2 系数应该为正。

多元回归分析也可以处理不同预测变量之间的交互作用。关于交互,我们是说一个预测变量对效标的效应取决于另一个预测变量的值。例如,假设西南部地区(有强烈的沙漠日晒)的人比东北部地区的人更在意皮肤癌的预防,这种情况下,与东北部地区相比,西南部地区的“风险”和“意图”两个变量之间的关系更强(即“风险”和“区域”之间存在交互作用)。在多元回归中,可通过设定一个交互项来反映变量之间的交互作用,这里在“风险”和“区域”之间设定了交互项。

$$\hat{Y} = b_1 \times \text{风险} + b_2 \times \text{区域} + b_3 \times \text{风险} \times \text{区域} + b_0$$

在上面的方程中,非零系数 b_3 表明“风险”和“区域”之间存在交互作用(此时“风险”和“区域”两个变量也包含在这个方程中)。

假设和困难

最常见的线性回归模型通常假设对预测变量的测量没有误差(固定效应模型【Fixed Effects Model】);预测变量中的误差会使得回归系数的估计有偏差。对于精确的统计推论来说,预测误差(残差)必须服从正态分布【Normal Distribution】。多重共线性【Multicollinearity】(即在预测变量之间存在较强的相关性)会导致回归系数很不稳定。回归分析对个别的极端值(游离值【Outliers】)非常敏感;单一个案就会显著改变回归系数的大小。回归诊断【Regression Diagnostics】可以估计出研究个案在多大程度上是强影响点【Influential Cases】,即能够改变回归系数甚至是回归方程总体拟合的个案。

历史和拓展

1900 年前后,在高尔顿和皮尔逊等人研究的基础上,多元回归方法开始作为一种方法进入科学文献中(详见回归分析【Regression Analysis】)。作为一种简单易行的技术,多元回归的发展离不开算法【Algorithms】的发展,以满足分析的数学处

理需要,尤其是在面对彼此相关的预测变量时。事实上,为了处理在应用多元回归分析时出现的计算难题,费希尔发展出一种正交因子方差分析【Analysis of Variance】法。随着大型计算机的广泛使用,多元回归分析在1960年代和1970年代得到了迅速发展。许多新的扩展随之涌现。广义线性模型【Genralized Linear Model】中包含了能够处理各种因变量的各种多元回归方法,包括处理二分变量或有序分类结果的Logistic回归【Logistic Regression】以及处理计数数据的泊松回归【Poisson Regression】。随机系数回归和多层次建模【Multilevel Modeling】可以对聚类数据(clustered data)进行分析,并在不同的聚类层次上对预测变量进行测量。非参数回归【Nonparametric Regression】、非线性回归和稳健【Robust】回归等模型具有很强的灵活性。

——Leona S. Aiken

(李洋译 高勇校)

*也可参见回归分析【Regression Analysis】。

参考文献

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). New York: John Wiley.

Fox, J. (1997). *Applied regression analysis, linear models, and related methods.* Thousand Oaks, CA: Sage.

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear regression models* (3rd ed.). Chicago: Irwin.

多重指标测量(Multiple-Indicator Measures)

对理论构念【Construct】进行测量【Measurement】是社会研究中最重要的步骤之一。把理论描述的抽象概念与这些概念【Concept】的经验指标【Indicator】联系起来,这对于清晰理解所研究的现象是至关重要的。事实上在社会研究中,在理论构造和经验指标之间建立联系,与在理论构念之间建立联系同等重要。

多重指标测量是指用不止一个的指标或题项来表示理论构念。与单一题项相比,多重题项测量有很多优势。第一,许多理论构念非常宽泛且复杂,难以用单独一个题项来表达。举一个简单的例子,没有人会认为仅仅通过美国政府考试中的一个"正确—错误"问题(true-false question)就能充分地测量一个学生对美国政府的了解程度。但是,如果针对这一主题有许多问题来提问,就可以全面地评估学生对这一主题的了解程度。

使用多重题项测量的第二个理由是准确性。单一题项缺乏准确性,因为它无法分辨某一属性的微细差别。事实上,如果题项是二分的,那么它就只能识别在这一属性上的两个层次。

采用多重题项测量的第三个原因是其信度【Reliability】较高。信度主要是指测量的一致性。采取相同问卷进行重复测量得到的结果相同吗?具有可比性的不同测验的结果是否相同?一般来说,多重指标测量包含较少的随机误差,而且与单一题项测量相比更为可靠,因为随机误差【Random

Error】在多重测量中被抵消了。基于上述原因,多重指标测量通常比单一指标测量能更好地表示理论构念。

——Edward G. Carmines
James Woods
(李洋译 高勇校)

参考文献

Carmines, E. G., & McIver, J. P. (1981). *Unidimensional scaling* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-024). Beverly Hills, CA: Sage.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Spector, P. E. (1992). *Summated rating scale construction: An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-082). Newbury Park, CA: Sage.

乘积(Multiplicative)

乘积一词是指一个变量乘以另一个变量的情况。这个词经常用在回归模型中,即因变量 Y 是变量 X 和变量 Z 的乘积的一个函数。一般表述为,Y 是 XZ 乘积的线性函数,即

$$Y = \alpha + \beta XZ$$

其中,XZ 是变量 X 乘以变量 Z。如果一个回归模型包括乘积项,那么,这个回归【Regression】模型就是"乘积的"。乘积项可以是一个预测变量【Predictor Variable】乘以另一个预测变量,或者预测变量乘以自身,比如

$$Y = \alpha + \beta X^2$$

存在两个不同变量的乘积项的模型【Model】,通常意味着模型中存在统计交互【Interaction】。如果模型中有表示二次、三次或更高阶的乘积项,通常表明 Y 和 X 之间有非线性【Nonlinear】关系。

上述乘积模型的拟合,要受到其预测变量如何度量的影响。X 和 Z 的取值方式(例如,从−5~+5,还是从 0~10)不同,得到的 Y 和 XZ 的相关【Correlation】也会不同。有很多乘积模型求值的办法,可以使乘积模型不受度量假设的影响。这些内容在安德森(Anderson, 1981)以及雅卡尔和图里西(Jaccard & Turrisi, 2003)的著作里中都有所阐述。

——James J. Jaccard
(李洋译 高勇校)

参考文献

Anderson, N. H. (1981). *Methods of information integration theory*. New York: Academic Press.

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression*. Thousand Oaks, CA: Sage.

多阶抽样(Multistage Sampling)

多阶抽样是指在调查【Survey】设计中,将总体【Population】单元分为不同层次,进

而在多个层次相应的多个阶段中进行抽样的抽样方法。在每个抽样阶段，只考虑那些在前一阶段选中的较高层次单元中的单元（不要与多相抽样混淆）。

最简单的多阶抽样是二阶抽样。在第一阶段，选中一个较高层次的单元。在第二阶段，对第一阶段中选中的较高层次单元下的较低层次的单元进行抽样。例如，第一阶段的单元是学校，第二阶段的单元是学生，或者第一阶段的单元是公司，第二阶段的单元是雇员。在对一个涵盖地理范围较大的家庭或个体的调查中（例如全国调查），第一阶段抽样单元可以是小的地理区域，包括数量适中的地址数或家庭数——大概 10 或者 20 个——在第二阶段抽样中，再从抽中的地理区域中抽取地址或家庭。

多阶段设计会使样本“聚集”在一个较高层次单元的有限组内。如果多阶抽样是要包括后面抽样阶段的所有单元——例如，每个被抽中的学校的全体学生——那么，这种样本就是整群抽样【Cluster Sampling】，因为它是由整个群体组成的。否则，如果在后面的阶段中继续进行抽样，那么这样的样本是群集样本（clustered sample）。

多阶抽样的主要目的是减少成本。在田野访谈调查中，样本中的每个群集都给访谈员带来了一定的工作量。因而，如果每个访谈员在同一个地点可以实施很多访谈，与单一阶段（未群集）抽样相比节省了路途的耗时。这就减少了访谈的单位成本，因此，在特定的预算下就能实现更大的样本规模。经常是这样（并不仅仅对于田野访谈调查而言），调查会有一个根据样本第一层次单元设定的固定成本。例如，在对学校学生的调查中，与每个学校建立关系是必需的，这就有一个成本，可能包括对学校的访问等，这部分成本与要进行抽样学生的数量无关。然后，通过把样本学生限定在数量有限的学校，也可减少单元数据采集成本。

但是，样本集中也有一个缺点。由于较高阶段单元中的抽样单元（例如，学校里的学生，较小区域的家庭）通常比总体中作为整体的一部分的单元更具同质性，因此，这种群集可能会降低调查估计的准确性。换句话说，样本群集会产生与样本比例分层相反的效应（参见分层抽样【Stratified Sampling】）。分层能确保所有的层次都能在样本中有所代表，群集使得只有一些类别的特征能在样本中反映出来。由此产生的准确性的降低可以通过群集的设计效应【Design Effect】进行测量。一般来说，每个群集层次的平均样本规模越大，群集单元的同质性越强，就越可能削弱抽样估计的准确性。例如，如果说一个班内的学生的同质性比一个学校的学生的同质性强，那么，从 100 个班级中每个班抽取的 10 名学生的抽样准确性较低，而从 100 个学校中，每个学校抽取 10 名学生的抽样准确性较高。

在设计调查抽样时，群集单元的选择以及给每个单元配置的样本量的选择都是很重要的。要根据支出和准确性两个方面要求来平衡。理想的做法是，在预算固定的情况下，由于能负担得起更大的样本规模所带来的准确度的提升，与由于群集本身造成的准确性的丢失相比，要尽可能地更大一些。总体的目标是在一定的预算条件下使得调查估计的准确性最高，或者说达到一个特定准确度的要求支出最小。

多阶抽样的另一个重要方面是被选概率的控制。如果在许多阶段后才抽出一个样本，那么一个单元的被选概率是每个阶段（条件）概率的乘积。例如，在一个特定学年对学校学生开展调查，二阶段抽样设计如下：选择学校，然后在每个学校中选择学生样本。那么，在学校 i 选择学生 j 的概率是 $P_{ij} = P_i \times P_{j|i}$，其中，$P_i$ 是选择学校 i 的概率，$P_{j|i}$ 是在选中学校 i 的条件下，又选中学生 j 的概率（很显然，可以推广到任何一个

抽样层次)。

在多阶抽样中,很重要的一点是从数据中计算出每个抽样阶段的被选概率。只有那样才可能进行无偏的【Unbiased】估计。此外,还有重要的一点是在设计阶段尽可能控制总体概率。尤其是在很多情况下,最有效率的设计是要赋予总体中各个单元相同的被选概率(见设计效应【Design Effect】)。但是,对于 P_i 和 $P_{j|i}$ 来说,有无数种组合取值为 P_{ij} 。一种解决方法是设每个概率都为常数。但这样做有一个缺点,就是不同群集单元的规模不同,因此,按比例抽取的样本规模大小也不同。这种差异之大,可能会导致实践和统计中出问题。因此,一个更常见的做法是,在除最后一个层次外的其他层次都采用"概率比例规模(probability proportianal to size,PPS)"抽样,然后在所抽中的前一层次的单元中,抽取最终层次所要求的单元数量。换句话说, $P_i = K_1 \times N_i$,且 $P_{j|i} = K_2/N_i$,其中, N_i 是在第 i 个第一层次单元中抽取的第二层次单元的数量(例如,在上一段的案例中,第 i 个学校中相关年级的学生数量), K_1 和 K_2 是常数。很容易发现,无论学校人数多少,这个研究设计都能保证相关年级的所有学生有相同的被抽中的总体概率 P_{ij},也使得每个抽中的学校都有相同的样本规模 K_2。

——Peter Lynn

(李洋译　高勇校)

参考文献

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York: John Wiley.

Scheaffer, R. L., Mendenhall, W., & Ott, L. (1990). *Elementary survey sampling*. Boston: PWS-Kent.

Stuart, A. (1984). *The ideas of sampling*. London: Griffin.

多策略研究(Multi Strategy Research)

莱德(Layder,1993)和布里曼(Bryman,2001)用多策略研究这个词来指将定量和定性研究相结合的研究。这一词汇要与多方法研究【Multimethod Research】区别开来。后者要求同时使用两种及以上方法,但并不要求必须是定量研究和定性研究相结合。多方法研究可以在定量研究或定性研究中同时使用两种及以上方法进行研究。相反,多策略研究认识到定量和定性研究是两种截然不同的研究策略,认为两种方法各有自己的认识论和本体论假设和标准。因而,多策略研究背后的观点是,这两种不同的研究策略并不是不可通约的研究范式【Paradigms】,而是可以出于不同原因、在不同情境中富有成效地结合起来的。

多策略研究有时会引发争议,因为一些作者认为由于定量研究和定性研究的认识论主张截然不同,因此,不可能真正结合起来(如 Smith & Heshusius,1986)。根据这种看法,那些看似结合了定量和定性研究的研究范例(例如,既进行半结构化访谈【Semistructured Interviewing】同时也进行结构化访谈【Structured Interview】为基础的问卷调查)并不能代表两种策略真正的整合,因为二者的基本原则是不可调和的。这只是在纯粹技术层面的一种表面性的互补而已。换言之,数据收集技术或分析技术可以结合,但是研究策略不可以结合。这种观点在定量和定性的争论中所赞同的立场是,两者只能围绕认识论而展开争论。

赞同定量和定性研究能够进行实质结合的作者和研究者认为,研究方法并不必然

嵌入认识论和本体论假设中。人们已提出了几种可以开展多策略研究的研究情景(Bryman,2001),包括:

(1)使用三角测量【Triangulation】,对定量和定性研究结果进行交叉确认(crosscheck);

(2)使用定性研究来辅助定量研究,反之相同(例如,用社会调查来帮助识别出那些适合采用定性工具如半结构化访谈进行访谈的人);

(3)使用一种策略来弥补使用另外一种策略的研究空白;

(4)使用两种策略来回答不同类型的研究问题;

(5)开展定性研究来帮助解释从定量研究中推导出来的结果;

(6)了解研究参与者的理解,表达研究者的关切。

在实践中,很难对这些多策略研究的不同情境进行区分。例如,迪肯、布里曼和芬顿(Deacon, Bryman, & Fenton, 1998)研究了英国大众媒体对社会科学的展现,他们采用各种研究方法来回答不同种类的研究问题。但是,他们发现当一些定量和定性研究结果冲突时,他们实际上进行了没有事先计划的三角测量。

——Alan Bryman

(李洋译 高勇校)

参考文献

Bryman, A. (2001). *Social research methods*. Oxford, UK: Oxford University Press.

Deacon, D., Bryman, A., & Fenton, N. (1998). Collision or collusion? A discussion of the unplanned triangulation of quantitative and qualitative research methods. *International Journal of Social Research Methodology*, *1*, 47-63.

Layder, D. (1993). *New strategies in social research.* Cambridge, UK: Polity.

Smith, J. K., & Heshusius, L. (1986). Closing down the conversation: The end of the quantitative-qualitative debate among educational enquirers. *Educational Researcher*, *15*, 4-12.

多元(Multivariate)

同时处理3个或更多变量的统计学就是多元的。例如,多重相关【Multiple Correlation】就是两个或更多个自变量与一个因变量的相关关系。在多元回归【Multiple Regression】中,至少有一个因变量和两个自变量。只要对列联表【Contingency Table】加入第三个变量,结果就变成多元的了。多元通常是相对二元【Bivariate】而说的,后者一次只考虑两个变量。虽然多个变量通常是自变量,但有时也可能是因变量。例如,在典型相关分析【Canonical Correlation Analysis, CCA】中,多个 X 变量与多个 Y 变量相关。分析人员也会对一个以上的因变量进行解释,如在联立方程【Simultaneous Equations】系统中就是如此。

——Michael S. Lewis-Beck

(李洋译 高勇校)

多元分析(Multivariate Analysis)

顾名思义,多元分析是用来对含有两个以上变量的数据进行分析的技术。有些技

术是新近才发展起来的，由于它们需要现代计算机的计算能力。也正因为多数技术都很新，它们没有统一的表达形式，也很难说对某一问题哪种技术最合适。

本词条提供了一个（非穷尽的）目录，以帮助研究人员根据数据类型或者统计问题来决定使用何种统计技术，同时也对每项技术进行了简要描述。本词条是根据要分析的数据数目编写的：有一个数据，还是有两个（及以上）数据。在两个数据的部分中，考察了两种情况：第一种情况是一个数据是预测变量（自变量）【Predictors (or Independent) Variables】，另一个数据是测量变量或因变量【Dependent Variables】。第二种情况是不同的数据集对应不同的因变量。

一个数据集

通常，要分析的数据表是由对一组单元（如被试）的多个测度组成。一般来说，单元是表格的行，测度是表格的列。

测量为定距或定比的：主成分分析

这种方法历史最久，用途最广。主成分分析（Principal Components Analysis，PCA）的目标是把数据中彼此相关的多个测度转换为一组彼此不相关的（即正交的）新变量。这些新变量在不同情境中被称为主成分【Principal Component】、因子【Factor】、特征向量【Eigenvector】、奇异向量、负荷。每个单元在每个主成分上都有一个赋分。

分析结果通常会用图表来展示不同单元在主成分上的投影（projections）以及不同变量的负荷（即所谓“相关圆圈”）。每个主成分的重要程度用其投影的方差（即特征值）或者解释的方差比例表示。在这种情况下，PCA 可以视为是对一个数据表的方差［也被称为惯量（inertia）］进行正交分解。

测量为定类或定序层次：对应分析（CA），多元对应分析（MCA）

对应分析（Correspondence Analysis，CA）是 PCA 在列联表【Contingency Tables】中的扩展。CA 的因子对列联表的卡方【Chi-square】进行正交分解。在对应分析中，表的行列是可互换的，可以表示为同一个图。如果有多个定类【Nominal】变量要分析，PCA 可以扩展为 MCA。CA 也被称为对偶尺度分析、最优度量【Optimal Scaling】或者互相平均法（reciprocal averaging）。

相似性或者距离：多维度量（MDS）、叠加树和聚类分析

当数据表的行和列都同样表示单元，测量的是单元之间的距离或者相似性时，就会用到这些技术。这类分析的目的是通过图示来直观表示单元之间的距离或者相似性。MDS 将不同单元在位图中用点表示出来，使得这些点的欧氏距离近似于初始相似性（original similarities）（经典的 MDS 等同于 PCA，用于测量距离，而非度量的 MDS 用于测量相似性）。累加树分析和聚类分析【Cluster Analysis】把各个单元比作一棵树的“树叶”，它们之间“在树上”的距离接近于原始距离或者相似性。

两个数据集，情况 1：一组自变量和一组因变量

多元线性回归分析

MLR 用多个自变量来预测一个因变量，自变量被假定是固定的或没有测量误差的，估计方法是最小二乘【Least Squares】。如果自变量之间彼此正交，问题就简化为一组双变量回归。如果自变量之间存在相关关系时，可以用偏相关系数来估计各个自变量的重要性。如果某个自变量可以被其他解释变量预测时，就会出现无法计算 MLR

的问题,这被称作多重共线性。对这一问题的解决办法会在下面进行讨论。

预测变量过多/多个因变量时的回归

偏最小二乘回归(PLSR)

PLSR 能够解决多重共线性问题,办法是计算同时解释自变量和因变量的潜在向量(与 PCA 中的成分类似)。这种技术用途广泛,在需要预测多个因变量时也可以使用。它兼有 PCA 和 MLR 的特别:可以像 PCA 一样把单元得分和变量负荷用制图的形式表现出来,也可以像 MLR 一样估计因变量(及其区间估计)。

主成分回归(PCR)

在 PCR 中,首先对自变量进行 PCA,然后用单元取值作为预测变量运行标准的 MLR。

岭回归(RR)

RR 通过在相关矩阵的对角线增加一个较小的常数项(岭)来解决多重共线性的问题。这样就可以计算出 MLR 的估计值。

降秩回归(RRR)或者冗余分析

在 RRR 中,首先对因变量进行 PCA,然后将单元取值作为一系列标准 MLR 回归中的因变量,而原初的自变量仍然作为预测变量(这个方法类似于反向的 PCR)。

多元方差分析(MANOVA)

在 MANOVA 中,自变量与标准方差分析中的自变量结构相同,但要预测多个因变量。MANOVA 计算出因变量的一系列有序正交线性组合(即因子),其中第一个因子纳入 ANOVA 时 *F* 值最大。*F* 值的抽样分布会进行调整以顾及其结构。

预测定类变量:判别分析(DA)

DA 的数学原理与 MANOVA 一样,它用一组自变量来预测给定单元属于哪个类别(即因变量为定类时)。判别分析合并因变量以使这些分组用于 ANOVA 中的固定因子时能够产生最大的 *F* 值。

拟合模型:验证性因子分析(CFA)

在 CFA 中,研究人员先是生成一个(或一些)潜在的解释结构模型(即构念),这个模型通常用图表达出来。因变量之间的相关性被拿来用这个结构拟合,然后会根据模型对数据的拟合程度来进行模型评估。CFA 的各种形式也被称为结构方程模型【Structural Equation Modeling, SEM】、LISREL 或 EQS。

两个(或以上)数据集,情况 2:两组(或以上)因变量

典型相关分析

典型相关分析(Canonical Correlation Analysis,CC)将因变量合并起来,以找到具有最强相关性的新变量对(称作典型变量[CV],每个数据表对应一个)。典型变量之间虽然具有最强的相关性,但并不一定能解释原始表格的大部分方差。这就使得有时对 CV 的解释很困难,但是 CC 依然是一个重要的理论工具,因为大多数多元技术都可以解释为 CC 的一个具体的例子。

多重因子分析(MFA)

多重因子分析(Multiple Factor Analysis, MFA)将几个不同的数据表合并成一项单独的分析。第一步是对每个表格进行 PCA。然后把表中所有单元格都除以主成分的第一个特征值,由此来把所有数据表标准化。这一转换类似于单变量的 *Z*-值,可以使每个表格在最终解中的权重相等,因此,可以同时分析多个异构数据表格(heterogeneous data tables)。

多重对应分析(MCA)

多重对应分析(Multiple Correspondence Analysis,MCA)可以用来分析多个列联表;它是对应分析的扩展。

Parafac 模型和 Tucker 3 模型

这些技术处理的是三维数据矩阵,用 PCA 的得分和负荷分解法来得到 3 个负荷矩阵(每个矩阵对应一个数据维度)。它们的不同之处在于对分解的约束条件不同(Tucker 3 模型生成的是正交负荷,而 Parafac 模型生成的不是)。

Indscal 模型

当每个被试都能生成一个单元和变量相同的数据矩阵时,就可以使用 Indscal 模型。Indscal 能生成一个共同的欧几里德解(及相应维度),并将被试之间的差别表示为在公共维度上的重要性的差别。

Statis

如果三维表格中至少一个维度是所有表格所共有的(如在多个时点用不同的变量对同样的单元进行测量),那么就可以使用 Statis 方法。这种方法的第一步是对每个表格进行 PCA,并对每个表格单元之间的相似性生成一个表格(即积差)。通过计算积差矩阵并进行 PCA(无须中心化),上述相似性表格被加以合并。分析中第一个成分的负荷被作为权重来计算折中数据表(compromise data table),即所有表格的加权平均数。最后将初始表格(及其单元)投影到折中空间(compromise space)中以探索其异同。

截长补短分析法(PA)

截长补短分析法(Procustean Analysis,PA)用于比较来自同一对象(objects)的距离表格(distance tables)。第一步是将表格转化为多维度量图,然后用截长补短分析法找出使得对象在两个图中的位置(在最小二乘的意义上)尽可能接近的一组转换。

——Hervé Abdi

(李洋译　高勇校)

参考文献

Borg, I., & Groenen, P. (1997). *Modern multidimensional scaling*. New York: Springer-Verlag.

Escofier, B., & Pagès, J. (1988). *Analyses factorielles multiples* [Multiple factor analysis]. Paris: Dunod.

Johnson, R. A., & Wichern, D.W. (2002). *Applied multivariate statistical analysis*. Upper Saddle River, NJ: Prentice Hall.

Naes, T., & Risvik, E. (Eds.). (1996). *Multivariate analysis of data in sensory science*. New York: Elsevier.

Weller, S. C., & Romney, A. K. (1990). *Metric scaling: Correspondence analysis*. Newbury Park, CA: Sage.

多元方差分析和协方差分析(Multivariate Analysis of Variance and Covariance)

这里的“多”指的是在一个分析中使用几个因变量【Dependent Variable】。在如多重回归【Multiple Regression】等用法中,“多”经常是指多个解释【Explanatory】变量。多元分析在心理学和生物学等领域都被成功地应用:在心理学测验中,每个测验题项就是一个单独的因变量;在生物学中,在一个特定的有机体上有多种观测值。

多元方差分析(MANOVA)和多元协方差分析(MANCOVA)是在一次分析中使用

同一些解释变量来对多个定量【Quantitative】因变量进行联合统计分析的模型,在 MANOVA 中,所有解释变量都是定类【Nominal】变量;在 MANCOVA 中,有些解释变量是定量的,有些则是定性的【Qualitative】(定类的)。如果所有解释变量都是定量变量,那么这些模型也可以扩展为回归个案。这 3 种方法是所谓的多元一般线性模型【General Linear Model】的特例。

在有多个因变量和一组解释变量的情况下,采用方差分析【Analysis of Variance】和协方差分析【Analysis of Covariance】可以逐个分析每个因变量和解释变量之间的关系。MANOVA 和 MANCOVA 的不同之处在于所有因变量在同一个分析中出现。当多个因变量有共同之处时,例如,它们是在一个心理学测试中不同题项的得分,这种方法就更好。这种多元方法还能解决对同一数据进行多次统计检验时,如何处理总体显著性水平的问题。

解释

举一个简单的例子,假设有一组关于同一组学生的数学和口语两门学科的学习能力测试(Scholastic Aptitude Test, SAT)的得分数据。不同性别和种族的 SAT 得分有差异吗？可以对每组 SAT 得分进行单独的方差分析【Analysis of Variance】,两个方程式如下所示:

数学 SAT=性别+种族+性别×种族+残差项
口语 SAT=性别+种族+性别×种族+残差项

每个分析都能得到平方和【Sums of Squares】与 p 值【p Values】,可以推断出两个解释变量及其统计交互【Statistical Interactions】对 SAT 得分的影响。

但是,这样的两个单独分析并没有利用两个解释变量本身是相关的这一事实。如果有某种方法能够将这一信息考虑进来,那么,分析就更有解释力,有可能发现在两个单独的分析中并不显著的结果。这些目标可以用一个多元方差分析来达到,而不是两个单独的方差分析。

历史发展

随着方差分析和协方差分析过渡到相应的多元分析,计算的复杂性极大增加。因此,直到 20 世纪后期出现了必要的统计软件,这些方法才得到很好的发展和应用。如今,所有主要的统计软件都能进行多元方差分析和多元协方差分析。而且很多软件手册都对这些方法及其结果进行了很好的解释。

应用

让我们考察两组 SAT 得分和一个解释变量,即性别。在两个单独的方差分析中,分别研究了数学和口语的得分有没有性别差异。但是,我们能够生成两个测试分数的散点图来显示两个因变量之间的关系。为了识别每个回答者的性别,可以用两种不同颜色的散点,这样就很容易看清两组数据是重合的还是有所不同的。我们就能看出对每个变量进行单独考察难以发现两组之间的差别。极端情况下,男生和女生的两项测量得分的均值都相同,但是在散点图中,却可以看到两组得分存在着差异。这只有采用多元方差分析而非两个独立的方差分析才能发现这些差异。

进行多元方差分析和多元协方差分析的第一步是进行虚无假设【Null Hypothesis】检验,即各组是否在不同的因变量上均有相同的均值。在上面的例子中,这意味着要检验女生和男生是否有相同的数学和口语得分,以及是否所有的种族有相同的数学和口语得分。如果这些虚无假设被拒绝,与单变量下的做法相同,下一步就是要找到哪组的

组均值不同。

一般来说,有4种检验总体假设有效性的方法。它们分别是 Wilks 的 Lambda(λ)统计量、Pillai-Bartlett trace 统计量、Roy 最大特征根统计量和 Hotelling-Lawley trace 统计量。针对同一个虚无假设,这4种方法可能得到不同的检验结果,如何选择取决于检验的稳健性和统计功效(statistical power)。Wilks 检验是最常用的统计量,如果满足适当的假设,其检验统计量的分布与 F 分布【F Distribution】较为接近。

假设条件

与其他统计方法一样,研究数据必须满足研究方法所需要的假定,得到的 p 值才有意义。数据必须是来自潜在总体的一个随机统计样本,观测之间是彼此独立的。因变量要服从多元正态分布【Normal Distribution】,即每个因变量都要服从正态分布。最后每个因变量必须满足方差齐性,在所有观测组中任意两个因变量之间的相关性必须相同。

——Gudmund R. Iversen

(李洋译　高勇校)

作者的提示:在万维网上搜索 MANOVA,可以了解到对这些方法的很多不错的介绍,以及使用这些方法的案例。

参考文献

Bray, J. H., & Maxwell, S. E. (1985). *Multivariate analysis of variance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-054). Beverly Hills, CA: Sage.

Velleman, P. F. (1988). *DataDesk version 6.0 statistics guide*. Ithaca, NY: Data Description, Inc.

N

$N(n)$

样本【Sample】量的符号。大多数社会科学研究是针对来自总体的样本而非总体【Population】本身来实施的。只要数据【Data】来自样本，通常就必须公布样本的大小，比如 $N=256$，或者 $n=256$（在实践中 N 和 n 没有区别，在文献中这两种用法都有）。此外，在分析中，所公布的 N 通常是排除了在分析变量上有缺失数据的个案。例如，一项调查采访了 1 000 个被调查者，但是对感兴趣的 6 个变量来说，只有 873 个个案没有缺失数据，因此有效的样本量 $N=873$，而非起初的 1 000。

——Michael S. Lewis-Beck

（李洋译　高勇校）

N6

N6（正规表达是 NUD * IST）与 NVivo 类似，是用于定性数据分析的专业计算机软件。更多信息参见相关网站。

——Tim Futing Liao

（李洋译　高勇校）

叙事分析（Narrative Analysis）

在人文科学中，叙事分析是指具对有故事形式的各种文本进行分析的一些方法。国家和政府会构建自己喜好的历史叙事，社会运动、各类组织、科学家、其他专业人士、民族/种族群体、用经验来编织故事的个体也都如此。如此多样化的文本都被称为“叙事”，是因为它们皆具有“顺序”和“后果”：事件经过了选择、组织、关联，然后被认为对某一受众是有意义的。讲故事者解释世界且体验世界；他们有时还创造关于“世界应该如何”的道德传说（moral tales）。叙事代表了进行认知和沟通的故事化方式（Hinchman & Hinchman，1997）。我在此着重讨论对个人经验进行口头叙事。

对叙事的研究兴趣萌生于一些当代运动：在人文科学中远离实证主义者【Positivism】研究方式和宏大叙事理论（如马克思主义）的“叙事转向”；在文学和流行

文化中的“回忆录热潮”；美国、欧洲和跨国运动中的“身份政治（identity politics）”，即有色人种、女性、同性恋者、其他边缘人群的解放斗争；迅速发展的疗愈文化（therapeutic culture），即在各种心理疗法中对个人生活的探索。“立足于普通人、边缘人、沉默者的生活，个人叙事回应了宏大叙事的瓦解，人们通过讲述与书写自身故事来理解经历、确立身份、重获新生”（Langellier，2001，p. 700）。

不同的研究者对个人叙事的定义有很大差异，这往往与学科背景有关。在社会史和人类学中，叙事指的是一个完整的生活故事，它是根据各种访谈、观察和文献编写而成的（如 Barbara Myerhoff 对加利福尼亚州威尼斯的犹太老人进行的民族志研究）。在社会语言学和其他领域中，叙事的概念被限定为围绕人物、场景和情节组织起来的简短的、有明确主题的故事［如回应单独一个访谈问题的拉波夫式叙事（Labovian narratives）］。在另一个传统中（心理学和社会学领域是常见的），个人叙事包括大段谈话——从一次或多次访谈中得来的在一定情境中的对生活的详尽叙述。调查者对叙述的不同定义使得其分析方法并不相同，但都要求调查者构建文本以便进一步分析，要选择并组织文档、完成田野笔记【Fieldnotes】，并且/或者选择访谈转录【Transcript】稿中的某些章节仔细研读。叙事并不能自己说话，也并不具有天然的优点；它们需要被加以阐释，才能成为社会研究中的资料。

叙事分析的几种模式

叙事分析有好几种分类方式（参见 Cortazzi，2001；Mishler，1995）。我要介绍的是用来描述当前众多方法的一种有启发性的尝试，它尤其适用于对个人经验的口头叙事（关于组织叙事，请参见 Boje，2001）。分类并不是为了评出优劣高低，但我会对每一类都提出我的问题。实践中不同的方法可以结合起来；他们并不互斥，如同所有分类一样，边界是模糊的。每一类我都会从健康和疾病领域中举几个例子，诚然它们都带有我的偏好。

主题分析

重点是文本的内容，讲述的内容（“what” is said）比讲述的方式（“how” it is said）重要，讲述的结果（the “told”）比讲述的过程（the “telling”）重要。这一路数背后的（未被明示的）语言哲学是：语言可以直接而明确无误地通向意义。如与扎根理论家【Grounded Theorists】一样，调查者会收集许多故事，从这些资料中归纳性地得出几组概念。对叙事按主题进行分类是通常用来呈现资料的办法，同时会有一些案例会被拿来作为例证。

加雷思·威廉森（Williams，1984）在其早期属于疾病叙事流派的论文中，讨论了个人如何叙事性地重新建构其病因来应对风湿性关节炎对身份认同的影响，这是一个将身体、疾病、自我和社会联结在一起的阐释过程。他分析了 30 个人对他们如何得病的叙述，从而确立了一种分类法，使用了三个案例来作为示例；这些叙述表明了叙事主题是多样性的，扩展了视慢性病为生活扰动（biographical disruption）的已有理论。他的访谈摘录（interview excerpts）的叙事形式是经典的、以时间为序排列的，但是他并没有对这种叙事的形式进行分析。

卡罗尔·凯恩（Cain，1991）在她对饮酒者互助协会（Alcoholics Anonymous）成员的身份获取的研究中更进一步，其中使用了观察和访谈的方法。在饮酒者互助协会的典型故事中，对饮酒有一些普遍的见解，新成员参加这个组织时会知晓这些见解；久而久

之,他们学会了将其生活中的事件和经验转变为一种模式化的生活故事,饮酒者互助协会的听众很容易识别出这些故事。她发现一种普遍性的文化故事,然后分析它如何塑造了组织成员"个人"对其饮酒生涯中的关键时刻这样一种事件的讲述。她的工作尝试性地考察了叙事结构,继而演变成对文本类型的分析。

主题法有助于将大量个案进行理论化——在不同的参与者和他们汇报的事件上发现共同的主题元素。由此可以建构一种分类方式来丰富正在形成的理论。由于兴趣在于言辞的内容,分析者会关注故事中显而易见的意义,由此来阐释故事究竟说的是什么。语言被视作一种资源,而非调查的议题。但是,这种方法是否在模仿客观主义者【Objectivist】的探究模式,即认为研究主题不受调查者的理论视角、兴趣和提问方式的影响?说话的背景——访谈的背景,更广泛的制度和文化话语的背景——通常被忽略了。读者必须假设,当许多叙事都被归为同一主题时,他们这些话都在表达同一个意思。但是,如何处理那些"含糊之辞"呢?如何处理那些不能归到任何一种类型中的"胡言乱语(deviant responses)"呢?还有,如何处理那些"言外之意(the unspoken)"呢?

结构分析

重点转向了讲述的过程(the "telling")和讲故事的方式(the way)。主题内容不可忽视,但形式同样需要关注:讲述者是如何选择特定的叙事工具(narrative devices)来使得故事有说服力的。与主题法不同,在此需要严肃地看待语言,除了相关的内容以外,语言也成了细致研究的对象。

叙事分析的第一种方法可以说是由威廉·拉博弗及其同事在30多年以前提出的,这种结构法分析了总体叙事中从句的功能——它所起到的沟通作用。拉博弗(Labov,1982)之后修改了这一方法以便考察对暴力的第一人称叙述,这种叙述是简短、主题集中、以时间为序的故事。但是他还是保留了叙事结构的基本部分:摘要(概要和/或故事的核心);方位(时间、地点、人物和情境);使事件复杂化的行动(事件顺序或者情节,通常有一个危机或者转折点);评价(叙事者从行动中退后以便评论其意义,进行情感沟通——叙事的"灵魂");问题解决(情节的结果);以及终曲(结束故事,将行动带回当下)。并非所有故事都要包括上述全部要素,且要素的顺序也并非一成不变。拉博弗的微观分析令人信服地展示了不当谈话如何导致暴力行动(在酒吧或在街道上)。从一个小的叙事语料库和戈夫曼之前的研究工作出发,他发展出了请求规则理论(a theory of the rules of requests),这一理论解释了不同叙事者所经历的不同场景中暴力突发的原因。

民族志诗学(ethnopoetic)的结构方法适用于较长的叙事,这种故事往往不采用典型的以时间为序的形式。在戴尔·海姆斯和其他人工作的基础上,詹姆斯·吉(Gee,1991)分析了因精神分裂症而入院的一位女性病人的言论,发现其富有艺术性且很有意义。叙事由不同片断组成(而非按时间),它可以分解成思想片段、诗的小节、诗的段落、基于叙事的讲述方式的其他部分。意义和阐释都受到了口头叙事特征的制约。吉提出了一种并非基于句子的话语单元理论。

由于结构法需要考察谈话的句法特色和韵律特色,因此它不适用于大批量分析,但是它在详细案例研究和比较多个叙事描述上非常有用。对几个案例进行微观分析,可以建构理论来将语言与意义勾连起来,如果像主题分析那样假定语言是意义的透明反映,这些内容就会被忽视。调查者必须根

据研究项目的重点来决定转录中需要保留多少细节。那些不熟悉社会语言学的人有可能无法读懂访谈摘录,这可能导致跨学科沟通的难度增加。与主题法类似,严格应用结构法可能会忽略叙事的语境,如历史因素、互动因素和制度因素。研究场景和社会关系会对叙事内容形成制约,形塑了某一故事的讲述方式。

互动分析

这里的重点是讲述者和倾听者的对话过程。经验叙事通常发生在一些特定的场所,如医院、社会服务机构和法院等,在这里讲述人和提问人共同参与会话。互动法并不忽视主题内容和叙事结构,但是它的兴趣在于讲故事是一个共同建构的过程(a process of coconstruction),在这个过程中,讲述者和听众进行合作来共同建构意义。个人经验的故事围绕着讲述者的生活世界,被插入到了问答过程中。这种方法需要文本转录中包括所有谈话中出现的参与者,如果互动中辅助语言特征(paralinguistic features)[④]也可以记录下来就更好了。

一些研究问题需要进行互动分析。杰克·克拉克和米什勒·埃利奥特(Clark & Mishler, 1992)试图讨论"体贴的"问诊(medical interviews)有哪些特征。通过分析停顿、打断、主题串联和谈话的其他方面,他们揭示了问诊如何能够(以及不能)让患者进行讲述,以便提供信息来进行准确的诊断和治疗。

苏珊·贝尔(Bell, 1999)比较了两个妇女在参与妇女健康运动前后的疾病叙事。她审视了自己在研究访谈中的参与情况,表明随之呈现的叙事是如何立足于具体的历史背景和政治影响当中。情境形塑了女性生活中的各种可能性、她们对疾病的体验、妇女与作者共同生成的某种疾病叙事。除了叙事组织和结构,对语言和互动的微观分析(microanalysis)在她的研究方法中也是至关重要的。

互动法对于研究各种场景下(法院、教室、社会服务机构、心理治疗办公室和研究访谈自身)讲述者之间的关系非常有用。与结构法类似,互动研究通常能够表现出言语的复杂性,而不只是把它视为简单的传达内容的工具。和在谈话分析【Conversation Analysis】中一样,文本转录可能对于缺乏相关经验的人来说很难。通常需要包括谈话中的停顿、结巴及其他方面,但是文本转录不同于录影带,它不能展现那些"言外之意(the unspoken)"。如何处理那些受访者的姿势、凝视和其他表现呢?

表演分析

表演分析是对互动法的进一步扩展,它不仅对言辞感兴趣,正如舞台隐喻所暗示的,讲故事被视为经历过往的"自我"的一种表演,讲述者(可能)通过语言和动作来感动听众,因此它不仅是一种"述(telling)",更是一种"行(doing)"。述行法中也有不同的形式,一端是拟剧论,另一端认为叙事是一种社会行动形式的叙事实践论(narrative as praxis)。因此,叙事研究者可以分析不同的特征:舞台上用口头叙事表达的演员(如故事中的角色及其自我定位是叙事者还是主角);场景(表演的条件以及表现故事的场景);角色之间的对话(引述的言辞);观众反应(随着剧情的展开而对戏剧进行理解的听众,以及在之后的阅读中进行解释的解释者)。尽管戈夫曼较早就提出了拟剧论(dramaturgic),但表演分析在叙事研究中是

---

④ 辅助语言特征,是指互动过程中除了正式语言之外的声调、表情、行动等特征。——译者注

新近出现的,研究者们在对身份(即"自我"的可行呈现)研究中用它做了种种尝试(Riessman,2003)。

克瑞斯汀·朗热利那和埃里克·彼得森(Langellier & Peterson,2003)提出了令人信服的理论和若干经验案例,包括对家庭(小组)故事讲述的细节分析、乳腺癌患者的疾病叙事等。他们分析了在每个表演中故事讲述者、听众和角色的定位;讲故事是沟通实践,它是具象的、置于特定背景的、有形的、散漫的、允许辩解也允许批判的。

述行论适用于对沟通实践的研究和对身份建构的详细研究——讲述者想要他人如何认识自己,他们如何将受众拉进来一起"构建"自己身份的。这种方法激发了人们去研究受众是如何参与到叙事表演当中的。正如 Wolfgang Iser 和读者反应论者(reader-response theorists)所提出的,读者是最终的阐释者,他们对叙事的理解也许与讲述者和研究者都不同。将视觉(通过摄像和摄影)叙事与口头叙事整合在一起,这种做法代表了当代的一个创造性转向(Radley & Taylor,2003)。

结论

叙事分析不再单单是文学研究的范畴;它已经渗透进了所有的人文科学和实践专业。上面回顾的多种方法适用于不同的研究项目和文本类型,但是每种方法都提供了一种系统的研究个人经验叙事的方式。批评者(有时合理地)认为叙事研究把内在的"自我"实体化了,貌似提供了一种"真实的"声音、一种纯粹的主观真实,将个体能动性理想化了(Atkinson & Silverman,1997;Bury,2001)。个人叙事过度个人化(overpersonalizing)的危险确实存在。

叙事法不适用于研究大量的无名无姓、面目不清的对象。有些分析方法既缓慢又辛苦,需要关注细微之处:言论的细微差别,回答的组织结构,研究者和研究对象的关系,以及社会和历史情境——使"个人"故事成为可能的文化叙事(cultural narratives)。在最近的反身性转向(reflexive turn)中,自传民族志【Autoethnography】和来自其他传统的学者正在产生他们自己的叙事,将自己的传记与研究素材联系起来(Riessman,2002)。

叙事并不是过去的镜像,而是过去的折射。想象和利益都会影响讲述者如何关联起各个事件,如何让事件对他人有意义。叙事对研究的用处,恰恰在于讲述者会对过去进行阐释而非复制过去。叙事记录的"真实性"并不在它忠实地再现了过去,而在于它在过去、现在和未来之间建立起了不断变化的联系。叙事是讲述者对生活进行重新想象的一种途径(如同叙事促进了国家、组织、民族/种族、其他群体形成了集体身份)。用赖特·米尔斯的话讲,叙事分析可以在个人生涯和社会结构之间(即私人性与政治性之间)建立起联系。

——Catherine Kohler Riessman

(李洋译 高勇校)

参考文献

Atkinson, P., & Silverman, D. (1997). Kundera's "immortality": The interview society and the invention of self. *Qualitative Inquiry*, *3*(3), 304-325.

Bell, S. E. (1999). Narratives and lives: Women's health politics and the diagnosis of cancer for DES daughters. *Narrative Inquiry*, *9*(2), 1-43.

Boje, D. M. (2001). *Narrative methods for organizational and communication research.* Thousand Oaks, CA: Sage.

Bury, M. (2001). Illness narratives: Fact or fiction? *Sociology of Health and Illness*, *23*(3), 263-285.

Cain, C. (1991). Personal stories: Identity

acquisition and self-understanding in Alcoholics Anonymous. *Ethos*, *19*, 210-253.

Clark, J. A., & Mishler, E. G. (1992). Attending to patients' stories: Reframing the clinical task. *Sociology of Health and Illness*, *14*, 344-370.

Cortazzi, M. (2001). Narrative analysis in ethnography. In P. Atkinson, A. Coffey, S. Delamont, J. Lofland, & L. Lofland (Eds.), *Handbook of ethnography*. Thousand Oaks, CA: Sage.

Gee, J. P. (1991). A linguistic approach to narrative. *Journal of Narrative and Life History*, *1*, 15-39.

Hinchman, L. P., & Hinchman, S. K. (Eds.). (1997). *Memory, identity, community: The idea of narrative in the human sciences*. Albany: State University of New York Press.

Labov, W. (1982). Speech actions and reactions in personal narrative. In D. Tannen (Ed.), *Analyzing discourse: Text and talk*. Washington, DC: Georgetown University Press.

Langellier, K. M. (2001). Personal narrative. In M. Jolly (Ed.), *Encyclopedia of life writing: Autobiographical and biographical forms* (Vol. 2). London: Fitzroy Dearborn.

Langellier, K. M., & Peterson, E. E. (2003). *Performing narrative: The communicative practice of storytelling*. Philadelphia: Temple University Press.

Mishler, E.G. (1995). Models of narrative analysis: A typology. *Journal of Narrative and Life History*, 5 (2), 87-123.

Radley, A., & Taylor, D. (2003). Remembering one's stay in hospital: A study in recovery, photography and forgetting. *Health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine*, *7*(2), 129-159.

Riessman, C. K. (2002). Doing justice: Positioning the interpreter in narrative work. In W. Patterson (Ed.), *Strategic narrative: New perspectives on the power of personal and cultural storytelling* (pp. 195-216). Lanham, MA: Lexington Books.

Riessman, C. K. (2003). Performing identities in illness narrative: Masculinity and multiple sclerosis. *Qualitative Research*, 3(1), 5-33.

Williams, G. (1984). The genesis of chronic illness: Narrative re-construction. *Sociology of Health & Illness*, *6*(2), 175-200.

叙事访谈(Narrative Interviewing)

叙事访谈与典型的深度访谈有何区别?前者要求生成关于经验的具体“故事”而非一般性描述。叙事可以有多种形式,从详细叙述具体历史事件(有清晰的开始、过程和结尾)的、边界严格的叙事,到穿越时间和地理空间的叙事——以整个生命或生涯为参考的传记式描述。

叙事访谈的出现代表着人文科学研究中访谈方法的一大转向。一问一答(刺激/响应)的模式让位于作为一种更发散性的工作的访谈。参与者加入到叙事的展开中来;叙事者和听众/提问者共同生产和建构所叙述的事件和经验的意义(Mishler,1986)。“促发性的(facilitating)”访谈者和作为信息承载者的(vessel-like)“被访者”,被两个共同生产出意义的积极的参与者替代(Gubrium & Holstein,2002)。与依靠开放式问题【Open-Ended Question】和/或封闭式问题【Closed-Ended Question】的主流社会科学访谈相比,叙事访谈与当代的民族志【Ethnography】共同之处更多。

鼓励叙事

如果把访谈看作是一种谈话,一种说话人之间的对话,那么日常谈话的原则就可以适用于此:轮流着说,要切题,进出故事世界时要有一些话头和话尾。一个故事能引发

另一个故事;当叙事者和提问者/听众为这样一些话题转换留出足够余地时,它有助于我们探索可能把几个故事联系起来的关联和意义。如果我们想要了解个体经验的全部复杂性,就要靠细节,即要了解具体的事件而非对经验的一般评价。与主流研究实践中的一般"自然"谈话相比,叙事描述(narrative accounts)在谈话中需要更多的铺垫。

为了把研究访谈开放为研究参与者的扩展性的叙事,研究调查人员就需要减少一些控制。尽管研究者可以遵循特定的经验路径,但是叙事访谈就意味着要追随研究参与者的足迹。对某个现象的真正的研究发现往往来源于访谈中的权力共享。

叙事往往在最意想不到的时间发生,甚至是在回答固定回应(是/否)问题时(Riessman,2002)。但是与其他问题相比,某些类型的开放式问题更有可能产生叙事的机会。比较一下这两个问题:"X何时发生?"和"告诉我发生了什么?然后发生了什么?"第一个问题是在询问一个孤立的信息片断,后者询问的是一些过往事件的扩展性描述。一些调查者在介绍情况之后就要求研究参与者讲述他们的故事——例如,疾病是怎么开始的。但是经验总是不止于对它的描述和叙述;事件可能被说得轻描淡写、无足轻重。只有通过进一步的追问才能促使参与者回忆细节、转折点以及认知中的其他转变、情感和行动。在我个人对期望生命历程的中断(如离婚)的研究中,我提出这样的问题:"您能记得……发生时的某一片刻吗?"我可能会进一步追问:"发生了什么事情使你记住了婚姻中的这一时刻?"研究卫生职业教育问题的科塔兹及其同事(Cortazzi, Jin, Wall, & Cavendish, 2001)提出:"在你最近的学习中是否取得了什么突破性进展?"通常他们会回答"啊,是的",然后是夹带着隐喻和情感流露的对这一突破性进展的长时间的叙述——一个学生将之称为"一声霹雳(a clap of thunder)"。

一般来说,访谈安排中的结构性越少,研究参与者(访谈者和受访者一样)就被给予了更多的控制权来用现有的文化形式共同建构叙事(construct narratives)。并非所有父母经常给子女讲故事,并非所有文化都建立在口述的基础上(当然在有些群体中,故事是讲述过往的主要方式)。研究访谈中的参与者使用讲故事来作为一种了解和倾诉的方式的程度是不同的。并非所有的叙事都是严格意义上的(社会语言学上的)"故事"。

收集和比较人们讲述过往的不同方式,这将会大有所获。一位叫贾尼丝·戈德曼的研究生研究了一组苏丹男孩,这些苏丹男孩跋涉了很多国家,最终被美国接受成为"合法的"难民。他们的叙事描述在很多重大方面差异很大,但是所有人在叙事中都会包括某个事件——例如穿越满是鳄鱼的、许多男孩葬身其中的河流。叙事中不带情感、不加评价,这与其他难民群体的叙述完全不同,这反映出了这些受访男孩的年纪较轻、他们与访谈者的关系以及文化和心理因素。

有些时候对于某个参与者来说,只让他们用口头语言进行叙事几乎是不可能的。民族志学者温迪·勒特雷尔(Luttrell,2003)在一个班级中针对怀孕青少年开展研究工作,她们中的大多数是非洲裔美国人,研究者期望每个女孩都能就一些关键事件进行叙述:得知怀孕、告诉母亲或者男朋友、决定保留孩子以及其他重要时刻。但是没人回答,为了打开叙事的世界,她鼓励女孩们进行艺术创作和角色扮演。当研究者要求女孩们讨论她们的艺术作品时,她们开始对那些关键时刻进行讲述——集体说故事。仅仅依靠从个体访谈中形成的文本,即我们"神圣的记录"来进行研究,这对调查者来

说是有局限的。当代学者在研究中进行的各种创新包括合并观察(combining observation)、长久关系以及与参加者的历时会话,甚至是叙事访谈的影像数据资料(例如对参与者的环境进行录像、参与者拍摄的照片、对其他人照片的反应)。

总之,叙事访谈并非一组技术,也并非一定是自然而然的。如果能在一些研究情境中创造性地使用这种方法,它就能够让我们在社会研究中建立起对话关系(dialogic relationships),实现更平等的沟通交流。

——Catherine Kohler Riessman

(李洋译　高勇校)

参考文献

Cortazzi, M., Jin, L., Wall, D., & Cavendish, S. (2001). Sharing learning through narrative communication. *International Journal of Language and Communication Disorders*, *36*, 252-257.

Gubrium, J. F., & Holstein, J. A. (2002). From the individual interview to the interview society. In J. F. Gubrium & J. A. Holdstein (Eds.), *Handbook of interview research: Context and method.* Thousand Oaks, CA: Sage.

Luttrell, W. (2003). *Pregnant bodies, fertile minds: Gender, race, and the schooling of pregnant teens.* New York: Routledge.

Mishler, E. G. (1986). *Research interviewing: Context and narrative.* Cambridge, MA: Harvard University Press.

Riessman, C. K. (2002). Positioning gender identity in narratives of infertility: South Indian women's lives in context. In M. C. Inhorn & F. van Balen (Eds.), *Infertility around the globe: New thinking on childlessness, gender, and reproductive technologies.* Berkeley: University of California Press.

本地研究(Native Research)

认识论【Epistemology】和科学中占据优势的实证主义【Positivism】一直强调,研究者的研究工作必须具有客观性【Objectivity】,必须采用可以解决效度【Validity】、信度【Reliability】和概推性问题的方法。土著化【Going Native】一词通常被归于布罗尼斯拉夫·马利诺夫斯基,他基于对新几内亚特罗布里恩(Trobriand)岛民进行的广泛研究而提出了这一词(Malinowski, 1922)。在马利诺夫斯基对人类学家和田野研究"对象"之间的关系的反思中,他提出,同传统上研究者应当与研究客体保持距离不同,人们反倒应当"理解当地人的观点,他与生活的关系,来认识他对他自己世界的想象"(p.290)。也就是说,置身陌生环境的科学家应当入乡随俗,用一种与他们所研究的人和文化更具互动性和反身性的方式,投入研究尝试中。

自马利诺夫斯基首次提出"土著化"概念之后的七十年,不同学科的研究者们一直关注的是如何接近和研究那些在性别、种族、阶级、民族或者其他特征上与研究者不同的群体和/或主题。但是在过去十年中,人类学领域中出现了"本地研究者""本乡研究者(indigenous)""局内研究者(insider)"等术语(Hayano, 1979; Kanuha, 2000; Malinowski, 1922; Narayan, 1993; Ohnuki-Tierney, 1984; Reed-Danahay, 1997),这些术语表示的是民族志学者研究他们自身所属的社会和身份群体,以及/或者研究他们个人直接或经常经历的社会问题。本土研究的例子包括女性研究者研究乳腺癌,

携带 HIV/AIDS 病毒的民族志学者组织 HIV 呈阳性的当事人进行焦点小组【Focus Group】,出生于广岛的科学家调查日本的原子弹幸存者。

研究者自己研究自己的好处包括:能接近之前未被研究或难以触及的研究人群,对"局内人"本性的深入了解提升了分析性诠释的可能性和水平,对那些只能诞生于特定文化场景中的体验【Lived Experience】的方法和资料有着复杂的理解。但是,本地研究也充满挑战。要做到在参与观察者【Participant Observation】和研究者的多元身份之间进行平衡,就需要研究者有独特的技术和理解,在情感上和理智上与资料之间保持距离,以便不含倾向性地进行分析。这会相当困难,尤其是当研究者的首要社会身份(primary social identities)与所研究的对象相似的时候。此外,评判本地研究者的一个核心标准不仅仅是事先对研究群体有所了解,他还要有能力在以研究者(局外人)行事之时仍然被接纳为研究对象的群体成员(局内人)。

解决本地研究的难题,需要对那些既有"局内人"的视角又具备"局外人"研究方法的本土研究者保持敏感。

——Valli Kalei Kanuha

(李洋译 高勇校)

参考文献

Hayano, D. (1979). Auto-ethnography: Paradigms, problems and prospects. *Human Organization*, *38* (1), 99-104.

Kanuha, V. (2000). "Being" native vs. "going native": The challenge of doing research as an insider. *Social Work*, *45*(5), 439-447.

Malinowski, B. (1922). *Argonauts of the Western Pacific*. New York: Holt, Rinehart and Winston.

Narayan, K. (1993). How native is the "native" anthropologist? *American Anthropologist*, *95*, 671-686.

Ohnuki-Tierney, E. (1984). "Native" anthropologists. *American Ethnologist*, *11*, 584-586.

Reed-Danahay, D. E. (Ed.). (1997). *Auto/ethnography: Rewriting the self and the social.* Oxford, UK: Berg.

自然实验(Natural Experiment)

自然实验或自然主义实验之所以是"自然"的,就在于这种方法调查的是现实生活,是自然发生的事件的展现,研究者不进行任何控制【Control】或者操纵,对研究结果通常也没有任何预设概念。在社会科学中,这种类型的经验研究在自然的场景中进行,例如操场、学校、工场、邻里、社区以及家庭。在这些背景中自然地发生了诸多事件,由此为社会科学家们提供了空间,他们用定量研究【Quantitative Research】和/或者定性研究【Qualitative Research】方法低调地收集并分析实验数据(例如深度访谈【In-Depth Interview】、个案研究【Case Study】、观察法【Observation】、问卷法【Questionnaire】、态度量表【Attitude Scale】、评级表和检验得分)。这类实验研究的是自然发生的现象对响应/因变量【Dependent Variable】的影响。

自然主义实验经常与调查事件受到研究者有意的系统性操控的实验室实验【Laboratory Experiment】或者田野实验【Field Experiment】相比较。例如,研究者会引入、排除或者操纵一些变量而将其他变量设为恒量。尽管存在这些差异,这两种方法是可以互相补充的,自然实验有助于确定精

确控制的实验室研究结果能否在现实生活中发生（即确立其外在效度【External Validity】分析）。另一方面，陈述性（indicative）的研究结果可以激发更为严谨的实验工作。自然实验的另一个优势是，有些研究只能通过自然实验的方法来实施，因为实验研究可能存在伦理、成本太高、难度太大等问题。

自然实验的潜在缺点包括抽样【Sampling】问题、推广到总体时有局限性、从数据中难以进行因果推断。实验研究可以随机抽样，但在自然实验中研究者无法控制抽样过程。因而，研究者不得不依赖"自选择（self-selection）"。例如，St. Helena 的研究（参见 Charlton, Gunter, & Hannan, 2002）对电视观众和非电视观众进行了比较。研究结果能否推论到其他时间、地点和人物，这也与抽样密切相关。只有当有证据表明样本是总体的代表性样本【Representative Sample】时，才能进行普遍推论。最后，自然实验很难进行因果推论（causal inferences），除非通过适当控制表明其他因素没有对结果产生影响。例如，库克·坎贝尔和佩拉基奥（Cook, Campbell, & Peracchio, 1990）讨论了四种一般类型的自然实验研究设计，只有一种实验可能——但也并不一定能够——进行因果推论。

查尔顿等人（Charlton et al., 2002）的研究是自然主义实验的一个例子。研究者对孩子们在其收看广播电视之前以及之后 5 年的情况进行监控，采用多元方法【Multimethod】研究（观察法、评估、自报告、问卷、焦点小组【Focus Group】讨论、日记【Diaries】和内容分析【Content Analysis】），研究利用了自然发生事件的优势来评估电视对观众的影响。

——Tony Charlton

（李洋译　高勇校）

参考文献

Charlton, T., Gunter, B., & Hannan, A. (2002). *Broadcast television effects in a remote community.* Hillsdale, NJ: Lawrence Erlbaum.

Cook, T. D., Campbell, D. T., & Peracchio, L. (1990). Quasi-experimentation. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organisational psychology* (2nd ed., Vol. 1, pp. 491-576). Chicago: Rand McNally.

自然主义（Naturalism）

自然主义通常包含三个含义。哲学自然主义（Philosophical naturalism）认为世界上存在一个所有事物都遵循的客观的、自然的秩序（参见 Papineau, 1993）。这一观点实际上否认了笛卡尔的二元论，后者假设物质和意识是分离的。哲学自然主义通常意味着唯物主义，就是说我们所体验的意识都能进行物质上的解释（尽管万物皆可还原为意识的观点与自然主义者的立场也是一致的）。哲学的自然主义只是一种形而上学的推理或思考，但是它是自然主义最重要的概念——科学自然主义（scientific naturalism）的基础。

科学自然主义

科学自然主义中包括广泛的理论视角。有人认为，自然的统一性（unity of nature）是指探索自然的方法是统一的，那就是我们当前所使用的科学方法。自然主义的极端观点则更为强硬，采取了物质主义

(physicalism)和还原论的极端形式。以此看来,所有现象都可以化约为物理属性,诸如理性和精神价值这样的现象只不过是一些附属现象(epiphenomena)。

自然科学家的立场多在上述两者之间,他们多少都是自然主义者,但是研究社会世界的人则不然,即使他们自称为社会"科学家"。如果自然主义的本体论假设应用到社会世界中,那么结论就是社会世界与物质世界是一致的,或者来源于物质世界(Williams,2000,p.49)。进一步,如果我们按这种排序来解释人类与世界的关系,那么科学方法经过适当改造后也可以用来研究社会世界。这一观点是社会科学中两种最突出的方法——实证主义【Positivism】和唯实论【Realism】的基础。

根据我们对世界作出何种假设以及从这些假设推论而来的方法,可以将自然科学和社会科学中的唯实论者分成不同群体,但是他们都相信自然科学中合理的东西——如进行解释和一般化——在社会科学中也同样合理。自然主义者并不否认社会世界(与物质世界相比)在向他人展现的方式上存在重要的差异,而且承认这会产生方法差异。但是,他们认为自然科学内部同样存在这样的差异。例如,物理学家使用实验方法,天文学家则使用被动的观察方法。

对自然主义的反对

在哲学和其他社会研究中都有许多反对自然主义的声音。

反理性主义者(antirationalists),如后现代主义者以及许多对科学进行社会研究的学者,反对的不是自然主义的本体论,而是科学本身。他们认为科学"只是一种对话或者一种社会组织形式"(Kincaid,1996,p.8),因此并不具有认识论特权。这使自然主义成为多余的,因为它无法证明自然秩序的存在。对反理性主义的反驳,不仅来自自然主义的捍卫者,而且来自科学方法本身(Williams,2000,Chap.4)。

第二种反对观点认为科学自然主义是一种同义反复。科学是我们了解自然秩序的唯一方法,那些称为自然的事物都是通过科学来发现的。对此有一些侧面的回应,认为自然主义并不依赖科学的任何特定历史形式,而只是可以为自然或社会难题给出最佳真理的那些科学形式。

把自然主义作为社会世界研究的基础,对此最重要的反驳是在方法论层面进行的。多数反驳意见对"自然的统一性"持怀疑论立场,认为社会世界是一种意识的产物并关注其如何运作。有人认为,由此而生成的反馈机制形成了一种不确定性,这种不确定性导致我们不可能找到规律性(regularity)来进行解释和一般化,故而不能采用物理学或生物学的方式来研究社会世界。这种观点直接导向了两种方法论立场,其中第一种方法论立场某种程度上是反方法论(antimethodology)的:(a)科学是发现世界真相的最可靠的方法,但遗憾的是,社会科学不可能是科学的,因此根本没有可靠路径可以获取社会世界的知识;(b)社会世界是可知的,但是必须从人文主义立场来研究,要强调的是对世界的理解而非解释。

对上述方法论上的反驳,也有许多回应。第一,自然主义和方法论是相互独立的,因为不同的自然科学和不同的社会科学(甚至是不同的研究问题)都需要不同的方法。大致而言,用解释性的抽样调查就不可能做好人类学,反过来用民族志方法也不能做好政治民意调查。第二,在许多复杂系统中可以进行解释和一般化,在许多社会领域中也同样可以。在物质世界的许多研究中(例如湍流、基因变异等),解释也只是或然性的(probabilistic)。第三,社会世界的某些

特性使得我们难以了解它们,但是物质世界中同样如此。

社会科学中关于自然主义的争论还有一个难点,就是人们常常混淆实证主义和自然主义。社会科学中批判实证主义的人,往往把实证主义对科学的特定概念混同于对社会世界进行的全部科学研究。这种混淆的问题不仅在于它是误导性的,而且在于它忽视了自然科学家虽然是自然主义者却也常常是反实证主义者(antipositivists)。

民族志中的自然主义

自然主义的第三个用途是在民族志研究上。这种用法与前面的、旧的用法没有丝毫联系,并且通常是那些反对科学自然主义的人文主义者在使用。被称为民族志自然主义(ethnographic naturalism)的这种观点认为,人们从意义中创造他们自己的社会现实,为了了解这些社会现实,研究者在进行研究时应当将他或她对那个现实中的干涉降低到最低点。这个立场主要与符号互动论【Symbolic Interactionism】相联系。

——Malcolm Williams

(李洋译　高勇校)

参考文献

Kincaid, H. (1996). *Philosophical foundations of the social sciences: Analyzing controversies in social research.* Cambridge, UK: Cambridge University Press.

Papineau, D. (1993). *Philosophical naturalism.* Oxford, UK: Basil Blackwell.

Williams, M. (2000). *Science and social science: An introduction.* London: Routledge.

自然主义研究(Naturalistic Inquiry)

自然主义研究是指研究在真实世界的场景中进行,研究者不试图操控感兴趣的现象(例如,群体、事件、方案、社区、关系或互动)。研究所感兴趣的现象是"自然"展开的,就在于它不是由研究者为研究而预先设定的过程,例如在实验室或其他可控的场景中进行研究。研究者在真实世界的场景中进行观察,人们在他们感到舒适的、熟悉的条件下根据开放式问题接受访谈。

埃贡·古巴(Guba,1978)在其关于自然主义研究的经典论文中,明确了两个对各种类型的科学研究进行描述的维度:(a)科学家为开展研究提前对一些研究现象进行控制的程度;(b)对输出结果加以限定的程度,即研究中使用预先设定的类别或变量来对现象进行描述的程度。他将自然主义研究立场定义为一种以发现为导向,把调查研究者对研究设定的操纵降低到最低程度,同时不预先对未来的研究结果进行预先限定的研究。自然主义的研究与受控的实验设计【Experimental Design】恰好相反,实验设计中调查研究人员通过操纵、改变或者设定外部影响为常量等方式,对研究条件进行控制,同时对受限的结果变量(outcome variables)进行测量。此外,自然主义的研究采用开放式的、类似会话的访谈形式,而其他方式则采用预设了回答选项的问卷进行调查。

在最简单的受控实验研究中,研究者在测验开始前和结束后两个时间点收集数据,然后在受限的标准化测量中对试验组和控制组进行比较。如此设计是要假设干预是单一的、可辨识的、分离的和可测量的。此外,如此设计还假设一旦进行干预,就要保

持干预的相对持续和不变。

例如在项目评估时，在有可能对项目的适应和改进进行限制以防止其干扰研究设计的严格性时，受控的实验评估设计才能发挥最优效果。相反，在研究方案容易产生变化和转向的真实世界中，自然主义的研究采用动态流程导向（dynamic process orientation）对一段时期内过程、方案或干涉的实际操作和影响进行记录，以此来取代受控实验所强调的固定干预/结果。评估者开始理解、记录研究项目参与者每天的现实情况，不去操纵、控制或者消除情境变量（situational variables）或者项目发展，而是接受一个不断变化的项目的现实复杂性。用以评估的资料包括对理解参与者经历有重要意义的、浮现出来的任何东西。

自然主义实验就是观察者在真实世界的变化期间在场并记录变化前和变化后的一个现象。自然实验可以包括两组之间的比较，一组经历了一些变化，另一组没有经历。之所以将这些研究称为自然主义的，是因为变化是由真实世界的参与者而非实验室中的研究员进行引导的。

尽管如此，两者的区别并非仅限于在真实世界中或者在实验室中那么简单；在多大程度上是自然主义设计，实际上是落在连续体（continuum）两端的渐变过程，一端是完全开放的田野工作，另一端是完全控制的实验室条件的连续统一体。而在这两个端点之间，研究者控制和操作的程度是可以不断改变的。例如，恰巧研究者的提问可能会成为干扰因素，影响事件的自然发展。或者，在进行方案评估时，成员反馈也可能成为影响事件自然发展的干扰因素。如果研究者想要将数据收集中的干扰因素降到最低，可以采用非干预性观察作为研究的策略。

在农业研究中采用的方法也能更形象地说明受控研究和自然主义研究的区别。在农业检验站，研究人员通过预先设定的测量和严格控制来不断对干涉进行细致改变，以此来引导田野实验的发展，比如在作物施肥研究中。相反，农业研究中的自然主义研究则主要包括观察、记录和研究农民们在他们自己的农场进行的“实验”，这些实验没有研究人员的引导或者介入。

人类学家在进行民族志田野工作时会采用自然主义研究的方法。社会学家采用自然主义研究对邻里和组织进行参与观察【Participant Observation】。方案评估者使用自然主义研究来观察团队如何实施一个项目。所有社会科学家通常都用自然主义的研究方法来研究社会现象和人类互动是如何在真实世界的场景中展开的。

自然展现的设计弹性

自然主义的研究设计难以在田野工作之前完全设定。尽管研究设计会确定一个起始的关注点、观察计划、起始的导引性访谈问题，但是研究所具有的自然主义和归纳属性使得研究之初就具体设定操作性变量、列出可检验的假设、完成问卷编写、预先决定抽样方法既不可能也不适宜。自然主义的设计会随着田野工作的开展而自然展现（Patton，2002）。设计的弹性来源于自然主义研究的开放性。

——Michael Quinn Patton

（李洋译　高勇校）

参考文献

Guba，E. G.（1978）. *Toward a methodology of naturalistic inquiry in educational evaluation.* CSE Monograph Series in Evaluation，No. 8. Los Angeles：Center for the Study of Evaluation，

University of California, Los Angeles.
Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage.

负二项分布(Negative Binomial Distribution)

概率【Probability】分布是指随机变量【Random Variable】的可能取值范围以及这些值出现的概率。如果随机变量 Y 的概率质量函数(probability mass function,PMF)具有如下形式,那么 Y 就服从负二项概率分布:

$$p(y) = \binom{y-1}{r-1} p^r q^{y-r},$$
$$y = r, r+1, r+2, \cdots, 0 \leqslant p \leqslant 1$$

其中,p 是成功概率(事件发生),q 是失败概率(事件没有发生),r 表示首次成功(或发生)。数据来自一个只有两种结果的过程:成功或失败(发生或未发生)。符合负二项分布的随机变量,其均值和方差如下:

$$\mu = E(Y) = \left(\frac{r}{p}\right), \ \sigma^2 = V(Y) = \frac{r(1-p)}{p^2}$$

要采取何种类型的回归【Regression】模型,以及用何种概率分类作为估计基础,这要取决于数据的特征。用来得到分布的准则要与数据的原理(特征)一致。如果数据是事件计数形式,数据就会呈现某些特征,如下限为0、观测值数目有限。此类数据不服从正态分布【Normal Distribution】。泊松回归【Poisson Regression】模型通常适用于此类数据。

但是,泊松分布的基本假设是数据是独立同分布的,有时这并不成立。负二项分布放宽了这些假设,扩展了泊松分布。如果你有理由相信观测值并不独立或并不是稳定发生的,最好就用负二项分布来进行估计。要检验泊松分布合适还是负二项分布合适,需要检验泊松回归是离散不足的(underdispersion)还是过度离散(overdispersion)的。如果作为因变量的计数变量的样本均值大于方差,就是过度离散。过度离散对于标准误和显著性统计的影响类似于 OLS 估计中异方差性【Heteroskedasticity】的影响。

举例来说,如果你研究有多少人参与抗议行动,负二项回归模型就很有用。一个人参与抗议的决策可能受到其他人的决策的影响,因此并不独立;学生政治组织的成员比不隶属于任何政治组织的人相比更可能参加抗议。这就违背了泊松分布的独立性假设。如果你研究左翼社会运动的成长,你会发现这违背了泊松分布的第二个假设。如果一个国家的政府从极权主义过渡到民主主义,你可能发现社会运动的发生率是变动的——专制政权比民主政权对社会运动的压制性更强。很明显,这并不只是在时间序列【Time-Series】案例中才会发生;如在横断面分析(cross-sectional analysis)中,不同国家的发生率也可能并不相同。

如果你认为负二项回归模型适用于你的数据,你一定要识别变化率的分布(如果你怀疑不一致性存在)。这一问题会带来很多困难,而且模型误设【Misspecification】会产生不准确的参数估计值【Parameter Estimation】。如果你没有理论基础来确定 λ,那么可以考虑使用广义事件计数模型

(generalized event count model)。

——Christine Mahoney

(李洋译　高勇校)

参考文献

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data*. Cambridge, UK: Cambridge University Press.

King, G. (1998). *Unifying political methodology: The likelihood theory of statistical inference*. Ann Arbor: University of Michigan Press.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.

不符案例(Negative Case)

在定性分析中,通常会把资料分组来构成各种模式(patterns)(也被视为构念),期待这些模式之间会有一些差异【Variation】。但是,通过有目的的搜索或者某种巧遇,我们可能会碰到无论如何广泛地界定这一构念都不能符合上述模式的案例。这个个案通常被称为“不符案例”,因为它与一般模式不符。例如,假设有人在研究照料老年痴呆症患者的人,研究中识别出来一个构念是“情感痛苦”(Khurana,1995)。访谈过的所有照料者都表现出了一定程度的情感痛苦。但是,随着收集的资料增加,研究者会偶遇到几乎没有情感痛苦的受访者。这些受访者的出现并不意味着要否定已有的一般模式,但是这样的个案能为“情感痛苦”的构念提供额外的启发。

通过与不符案例的比较,分析者可以更清晰地解释概念【Concept】的属性是什么以及不是什么(即确定概念的边界)。不符案例有助于弄清楚某一概念成立的重要条件,例如它能让研究人员解释为何照料者或多或少都会有情感痛苦。此外,不符案例有助于研究人员扩展或者修正最初的构念,可能由此增加其解释力。不符案例还可以增强研究的效度【Validity】,表明分析者愿意考虑其他可能性并已经在寻求其他可能解释(Miles & Huberman,1994)。社会科学研究是在与人打交道,因此发现的每种模式总会有例外(Patton,1990)。

做研究时是否总能发现不符案例呢?答案是,能否发现不符案例取决于研究者付出多大的努力和多少时间来观察。研究者要用多长时间来寻求“不符案例”,对此没有固定的法则(Glaser & Strauss,1967)。一项研究项目总是在时间、经费和接触总体的渠道等方面有限制的。但是,研究者应当努力寻求模式的特例,在写作时要包括对这一寻求过程的叙述;如果发现有不符案例,应当在对研究发现的讨论中思考其他的思路和解释。这不仅仅会增加研究的可信性,而且加深了对调查现象的理解,对研究参与者和研究成果的使用者都有好处。

——Juliet M. Corbin

(李洋译　高勇校)

参考文献

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory*. Chicago: Aldine.

Khurana, B. (1995). *The older spouse caregiver: Paradox and pain of Alzheimer's disease*. Unpublished doctoral dissertation, Center for Psychological Studies, Albany, CA.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.

嵌套设计(Nested Design)

嵌套设计是这样一种研究设计:某个因子(如因子 B)的不同水平可以归到(或嵌套在)另一个因子(如因子 A)的不同水平中。由此,在一个嵌套设计中不可能对水平 A 和水平 B 的完整组合进行评估。例如,有人希望研究因特网搜索策略(因素 A)对大学生信息查找效率(因变量)的影响。随机选取一所大学中的 6 个新生英语班级;3 个班级被指派进行"线性搜索",另外 3 个班级被指派进行"非线性搜索"。表 1 显示了相关研究设计。

表 1

		因子 A:网络搜索策略		
		处理 1:线性搜索策略	处理 2:非线性搜索策略	行均值
因子B:新生英语班级	班级 1	$\bar{Y}_{1(1)}$		$\bar{Y}_{.1}$
	班级 2	$\bar{Y}_{2(1)}$		$\bar{Y}_{.2}$
	班级 3	$\bar{Y}_{3(1)}$		$\bar{Y}_{.3}$
	班级 4		$\bar{Y}_{4(2)}$	$\bar{Y}_{.4}$
	班级 5		$\bar{Y}_{5(2)}$	$\bar{Y}_{.5}$
	班级 6		$\bar{Y}_{6(2)}$	$\bar{Y}_{.6}$
	列均值	$\bar{Y}_{1.}$	$\bar{Y}_{2.}$	总均值 $=\bar{Y}_{...}$

录取到这些班级里的新生形成了"完整的组",他们无法以个人为基础被随机指派到两种处理条件下。此外,他们的学习过程和行为可能相互影响;学生信息查找行为在不同班级间的差异是嵌入每种处理条件中的。这个限制使得此设计成为一个嵌套设计而非完全交互设计,嵌套设计用 $B(A)$ 表示。

上述嵌套设计的假设统计模型是

$$Y_{ijk}=\mu+\alpha_j+\beta_{k(j)}+\varepsilon_{i(jk)},$$
$$(i=1,\cdots,n;j=1,\cdots,p;$$
$$k=1,\cdots,q)$$

其中:

Y_{ijk} 是第 i 个观测值在因子 A 的第 j 个水平上和因子 B 的第 k 个水平上的取值。

μ 是总均值(grand mean),是观测总体的一个常数。

α_j 是因子 A 的第 j 种处理的效应;从数学上来看,它等于均值(μ,j)对总均值(μ)的偏差。对于第 j 种处理条件下的所有观测值的因变量取值来说,它都是常数,此外还要满足不同处理条件下的 α_j 加总为 0 的约束条件。

$\beta_{k(j)}$ 是因子 B 的第 k 种处理条件的效应,嵌套在因子 A 的第 j 个水平中。从数学上来看,它等于同时处于第 k 种和第 j 种水平下时的均值($\mu_{jk.}$)与总均值(μ)的偏差。第 k 种处理条件下嵌套于因子 A 的第 j 种条件,对于这些观测值的因变量取值来说,它都是常数。该效应被假定在总体中服从正态分布。

$\varepsilon_{i(jk)}$ 是与因子 A 在第 j 种条件且因子 B 在第 k 种条件下的第 i 个观测值相关联的随机误差效应(random error effect)。它是一个在总体中服从正态分布的随机变量,独立于 $\beta_{k(j)}$。

这个案例是一个平衡的完全随机的嵌套设计。每个单元格(即水平 A 和 B 的组合)都有 $n=4$ 个观测值;因此设计是平衡的。由于观测值被随机归入 A 的不同水平(或条件)中,因此这个案例是完全随机的,在案例中也没有出现区组变量(blocking variable)。如果研究者控制个体差异在区组变量(如"以往进行网络搜索的经验")上的差异,那么就应当使用完全随机区组嵌套设计(completely randomized block nested design)。表2解释了这个设计的构架,以及它与完全随机嵌套设计的差异。

表 2

		因子 A:网络搜索策略			行均值
		因子 C:以往使用网络的经验	处理1:线性搜索策略	处理2:非线性搜索策略	
因子 B:新生英语班级	班级1	高	$\bar{Y}_{1(1)1}$		$\bar{Y}_{.1.}$
		低	$\bar{Y}_{1(1)2}$		
	班级2	高	$\bar{Y}_{2(1)1}$		$\bar{Y}_{.2.}$
		低	$\bar{Y}_{2(1)2}$		
	班级3	高	$\bar{Y}_{3(1)1}$		$\bar{Y}_{.3.}$
		低	$\bar{Y}_{3(1)2}$		
	班级4	高		$\bar{Y}_{4(2)1}$	$\bar{Y}_{.4.}$
		低		$\bar{Y}_{4(2)2}$	
	班级5	高		$\bar{Y}_{5(2)1}$	$\bar{Y}_{.5.}$
		低		$\bar{Y}_{5(2)2}$	
	班级6	高		$\bar{Y}_{6(2)1}$	$\bar{Y}_{.6.}$
		低		$\bar{Y}_{6(2)2}$	
	列均值		$\bar{Y}_{1..}$	$\bar{Y}_{2..}$	总均值 $=\bar{Y}_{...}$

除了平衡完全随机嵌套设计外,还有其他设计方式,如不平衡嵌套设计(unbalanced nested design)、平衡或不平衡随机区组嵌套设计(balanced or unbalanced randomized block nested design)、平衡或不平衡随机因子嵌套设计(balanced or unbalanced randomized factorial nested design)、平衡或不平衡随机区组因子嵌套设计。关于这些设计的细节,读者可以参考柯克(Kirk,1995),以及马克斯韦尔和德莱尼(Maxwell & Delaney, 1990)。

嵌套设计也可以被称作分层设计(hierarchical design),它多用于研究者对随机指派过程不进行控制的准实验【Quasi-Experiment】研究中。有时在课程研究、临床医学、社会学、行为学的研究中,观测对

象或分析单位是完整的一组对象(比如班级、治疗群体、帮派和牢房等),不能把这些组打散来把个体随机指派到不同处理条件中去,这时嵌套设计就是通行的,有时甚至是必需的。嵌套设计中观测单元之间并不是独立的,因此方法学家曾经一度建议使用组均值(如上面例子中的班级均值)进行统计分析。但是现在直接基于个体数据和相应统计模型的其他方法已经被证明是更优越的(Hopkins,1982)。处理观测单元之间互依性(interdependence)的更好方法,是使用如层级线性模型(HLM)这样高效的统计建模技术。在一个典型的关于教育的研究设计中,学生在班级中接受教育,班级嵌套于学校,学校嵌套于学区。HLM可以在三个水平进行建模:学生是水平1,班级是水平2,学校是水平3。每个较高水平(如水平2)模型都把在较低水平(如水平1)上所收集数据的嵌套特征考虑进去,因而HLM完整而充分发挥便抓住了多层数据的本质特点(Raudenbush & Bryk,1988)。

——Chao-Ying Joanne Peng

(李洋译　高勇校)

*也可参见横截面数据设计【Cross-Sectional Design】。

参考文献

Hopkins, K. D. (1982). The unit of analysis: Group means versus individual observations. *American Educational Research Journal*, *19*(1), 5-18.

Huck, S. (2000). *Reading statistics and research* (3rd ed.). New York: Addison-Wesley Longman.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Belmont, CA: Brooks/Cole.

Kirk, R. E. (1999). *Statistics: An introduction* (4th ed.). Orlando, FL: Harcourt Brace.

Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparison perspective*. Belmont, CA: Wadsworth.

Raudenbush, S. W., & Bryk, A. S. (1988). Methodological advances in analyzing the effects of schools and classrooms on student learning. *Review of Research in Education*, *15*, 423-476.

网络分析(Network Analysis)

网络分析是对社会关系的跨学科研究,萌生于人类学、社会学、心理学以及应用数学中。它从关系的角度来看待社会结构,其最基本的概念是社会网,社会网由位于社会最基本的层次上的社会行动者和联系这些行动者之间的关系连带组成。其基本的假设是社会行动者是相互依存的,他们之间的关系连带会对每个社会行动者以及由社会行动者所组成的社会团体产生重要的影响。

网络成员或节点可能是团体、组织或者人群。网络分析包括了相关的理论分析、模型建构和实证研究,还有(可能)复杂的数据分析技术。它的目的是研究网络结构,通常采用密度、中心性(centrality)、声望、互动和角色等概念来对网络结构进行分析。社会网数据组可能是多维度数据和/或纵贯数据,通常包括行动者属性等信息,如行动者年龄、性别、种族、态度和信仰。

社会网范式的基本预设是,单单依靠对行动者属性的了解进行解释是不够的,还需要了解其社会关系结构。如果所研究的个体行动者所处的社会情境是重要的,关系信息就应该被收集并加以研究。网络分析不是对个体进行测量,而是对关系连带(ties)

模式的数据进行分析,并审视关系连带的存在与运作如何受到个体行动者所嵌入的社会网的限制。例如,人们可能会测量一组工人的“沟通”关系、“邻里”关系、“敌意”关系、“社会支持”关系。有的网络分析纵贯研究,也就是将变化中的社会结构看作潜在过程的结果。其他的分析将个体与事件连接起来(即隶属网络,affiliation network),例如参与一系列社区活动的一组个体。

网络结构可以在多个层次上进行研究:二人组(dyad)、三人组(triad)、子群体甚至整个网络。此外,可以在多个不同层次上提出网络理论。尽管网络分析的多层次特征决定了它可以提出不同的结构问题并同时进行研究,但是它通常要求不能把每个个体视作独立的分析单位。对于完全网络或者整体网研究尤其如此,在这些研究中要对一个界定清晰的社会行动者总体进行普查,总体中不同行动者之间的各类关系连带都要进行测量。此类分析内容可能包括小群体中的结构平衡、通过无向连带(indirect ties)传递的信息流动、组织中的结构等价(structural equivalence)、多个组织之间的关系模式等。

例如,网络分析可以让研究人员建构组织成员的互依模型。这种分析范式有一系列相关概念、理论、方法,来调查在组织成员与工作相关的行动流程中以及在组织成员的不断演化的认知和信仰中,非正式的组织结构如何与正式科层结构互动交会。因此,它对诸如领导力、态度、工作角色、离职以及计算机支持的协同工作等多个组织行为议题都有所贡献。

历史背景

网络分析的发展起源于多个研究传统,包括(a)1930年代,由精神科医生雅各布·莫雷诺创建的社会测量学(sociometry);(b)1950年代和1960年代,致力于理解从部落村庄向多族裔城市的移民过程的民族志研究,尤其是拉德克利夫·布朗的研究;(c)发端于1950年代描述个人社群(personal communities)、社会支持(social support)和社会动员等性质的调查研究;(d)用于理解组织之间和国家之间关系结构的档案分析(archival analysis)。此外值得关注的还有列维-斯特劳斯的著作,他第一次对亲属进行了形式化界定,由此发展出了对关系的数理代数理论。此外,阿纳托尔·拉波波特第一个提出了经不同节点(nodes)的关系连带和关系流的精确统计模型。

在1940年代和1950年代,网络分析的发展重点是复杂数学模型的应用,尤其是离散数学和图论(graph theory)。在1960年代和1970年代,詹姆斯·戴维斯、塞缪尔·莱因哈德、保罗·霍兰、哈里森·怀特、马克·格兰诺威特和林顿·弗里曼在网络研究中提出了传递性、结构等价、弱关系的强度(the strength of weak ties)和中心性等概念。开始时许多研究各自为政,但到1970年代由于图论的提出和计算技术的发展,这一领域得以逐渐发展并交会融合。网络分析作为一个独立的研究领域,诞生于1970年代早期。诞生之时的重要事件包括哈拉里、诺曼和卡特赖(Harary, Norman and Cartwright, 1965)等人的开创性著作;在1970年代后期主要起源于加利福尼亚大学尔湾分校的网络分析软件的出现以及网络分析者年度大会,这一大会目前由“社会网络分析国际网络(International Network for Social Network Analysis)”主办。这些著名的社会网络大会又被称为“阳光地带(Sunbelt)”,每次能吸引多达400位的来自全球的与会者。许多领域(如组织科学)通过采取网络视角已经取得了快速发展。

多年来,社会网分析视角已经被越来越

多地用于理解社会科学和行为科学中的许多现象，包括（来自 Wasserman & Faust，1994）：

- 职业流动
- 城市化
- 世界政治和经济体系
- 社区精英决策
- 社会支持
- 社区心理学
- 小组问题解决
- 信息的传播和采纳
- 公司连锁（corporate interlocking）
- 信仰体系
- 社会认同
- 市场
- 科学社会学
- 交换和权力
- 舆论和社会影响
- 结盟

此外，它还可以用来理解许多当代问题，包括：

- 互联网
- 知识和分布式智能（distributed intelligence）
- 网络传播研究
- 恐怖主义
- 新陈代谢系统
- 健康、疾病和流行病学，尤其是艾滋病病毒

在对不同的网络研究方法进行详细论述之前，我们先了解一些重要的测量方法。

测量

完全网络

在完全网络研究中，要对预先设定的网络成员总体的网络连带进行一次普查。观察网络连带可以采用多种方法（如问卷调查、档案分析和参与观察等），观察的网络连带也可以是不同类型的。如果需要根据网络成员在更广泛的社会系统中的位置（如他们在网络中的中心性，或者更一般地讲，他们与其他网络成员相联系的模式）来理解其行为，那么完全网络研究通常就是适宜的。同样，如果感兴趣的是作为一个整体的网络的特征时，也需要对完全网络进行观察（如集中度、片段性或连通性）。

自我中心网络

完全网络研究在范围和规模上有所限制，因此一般不能够对那些大规模的，甚至无限的总体中的所有连带和节点进行研究。为了对这些现象进行研究，研究人员通常要使用问卷调查研究方法，对个体网络（经常称之为自我中心或者本地网络）样本进行研究。这些较小的网络包括连接位于网络中心的焦点个人（或自我）和与其关系紧密的“联系人”或“伙伴（alters）”的特定连带。这样的研究关注的是自我的连带，以及自我的伙伴（ego's alters）之间的连带。自我中心网络可能包括诸如直系亲属、弱连带、频繁联系以及提供精神和物质帮助等关系。这些关系的特征包括其多样性、内容、强度和结构。因此，分析者可以研究网络成员构成（如提供社会或情感支持的女性比例，或者说更一般意义上的基本行动者特征）、网络特征（如两人之间共有关系的比例）、关系关联量度（直系亲属的强连带关系是否意味着支持性关系？）网络结构（各种关系的结有多密集？行动者以任何有意义的方式聚集吗？）。

滚雪球抽样和连接追踪研究

另一种研究大网络的方法是对节点和连带进行简单的抽样。针对网络研究的抽样理论中有一些重要的研究成果（如对子图或者子分量的估计，许多方法来源于奥韦·弗兰克）和很多独特的技术或者策略，例如

滚雪球抽样,即在一个多阶段抽样过程中,先抽取一些节点,然后再抽取与这些样本相关联的样本,并重复此过程。在连接追踪抽样设计中,重点是在连接而非行动者——一组社会连接背后是从一个受访者到另一个受访者的关系。对于难以接近的或者是隐藏的总体来说,这种设计被看作获得样本节点的最实际的方法。

认知社会结构

关于社会认知的社会网研究,调查的是个体网络行动者如何认识他人的连带关系及身在其中的社会结构。这样的研究通常要求对一个网络进行多个视角的测量,例如观察每个网络成员对各成员之间关系连带的看法。大卫·克拉克哈特将所得到的数据阵列称为认知社会结构。研究的重点在于厘清社会认知与网络位置相关联的各种方式:(a)人们在社会结构中的位置可能会决定他们所能得到的特定信息,进而决定他们的感知;(b)结构位置可能与社会互动的特定模式相关;(c)结构位置可以通过影响人们对社会场所(social locales)的洞察来形塑社会认知。

方法

社会网分析人员发明了多种方法和工具来对关系数据进行研究。这些技术包括由数学家发明的图论法(其中许多要对不同类型的子图进行计数);由定量的社会学家和心理学家推广起来的代数模型;以及各种统计模型,这其中包括社会心理学的社会关系模型(social relations model)以及最近的随机图模型,后者是由奥韦·弗兰克和大卫·斯特劳斯首先在网络研究文献中进行介绍的。对上述模型进行拟合的软件包已经得到广泛应用。

最近在统计领域中网络研究方法出现了令人欣喜的发展,这反映了在社会科学和行为科学中对动态的基于网络的社会场景中社会行动者的互赖性问题的理论关注日益增长。因而,为结构性网络现象及其变迁建构理论和经验上可行的参数模型越来越重要了。统计计算中的巨大进展也使得研究者能够轻松地拟合复杂模型。

标记方法

在最简单的情况下,网络研究需要对代表个体行动者的一组节点 $N=\{1,2,\cdots,n\}$ 中的所有节点对测量是否存在有向或无向连带。观测到的将节点 i 和节点 j $(i, j\in N)$ 连接起来的连带可以用 x_{ij} 表示,如果连带存在,则观测连带取值为 1,否则取值为 0。网络可能是有向的(此时 x_{ij} 和 x_{ji} 是有区别的,取值可以不同)或者是无向的(此时 x_{ij} 和 x_{ji} 没有区别,取值相同)。其他情况包括:

1. 赋值网络,其中 x_{ij} 取值范围是 $\{0,1,\cdots,C-1\}$。

2.时间相依网络,其中 x_{ijt} 表示在某个时点 t 上,节点 i 到节点 j 的连带。

3.多重关系或者多变量网络,其中 x_{ijk} 表示从节点 i 到节点 j 的 k 类连带(一个固定的连带类型组,$k\in R=\{1,2,\cdots,r\}$)。

在大多数关于网络方法的统计文献中,集合 N 都被看作固定的,而网络连带被认为是随机的。在这个例子中,将节点 i 和节点 j 连接起来的连带可以用随机变量 X_{ij} 来表示,那么随机变量的 $n\times n$ 的数组 $\boldsymbol{X}=[X_{ij}]$ 可以看作关于 N 的随机(有向)图的邻接矩阵(adjacency matrix)。这些数组的所有可能的状态空间是 $\boldsymbol{\Omega}_n$。数组 $\boldsymbol{x}=[x_{ij}]$ 表示 $\boldsymbol{X}$ 的一个实现(a realization)。

图论技术

图论在网络分析的发展中扮演了重要

的角色。图论技术蕴含了理解网络中的凝聚性、连通性和碎片性的方法。网络的基本测量包括它的密度(可能连带在实际观测网络中的比例)及其节点的度序列(degree sequence of its nodes)。在一个无向网络中,节点 i 的度 d_i 是与节点 i 相关联的不同节点的个数($d_i = \sum_{j \in N} x_{ij}$)。在有向网络中,研究感兴趣的是点入度($\sum_{j \in N} x_{ji}$)和点出度($\sum_{j \in N} x_{ij}$)的序列。在网络中,对内聚子集(cohesive subset)进行描述和识别的方法取决于小集团(网络节点的子图,且节点成对相关联)的概念以及不同的概括方法(包括 k-小集团、k-丛、k-核心、LS-组和 k-关联子图)。

我们对连通性(connectedness)、连通度(connectivity)和集中度(centralization)的理解,也是通过网络中路径长度的分布建构起来的。从节点 i 到另一个节点 j 的长度为 k 的路径,可以用不同节点的一个序列 $i=i_1$, i_2,⋯, $i_{k+1}=j$ 来界定,其中的 i_h 和 i_{h+1} 是通过一个网络连带联结起来的。如果从 i 到 j 没有 $n-1$ 或更短长度的路径,那么 j 对于 i 来说就是不可达的,从 i 到 j 的距离就是无限的;否则,从 i 到 j 的长度就是从 i 到 j 最短路径的长度。在有向网络中,如果每个节点可以从其他任意节点到达,那么这个有向网络就是强连通(strongly connected)的;如果对于任何一对节点,至少从其中一个节点出发可以到达另一个节点,那么这个有向网络就是弱关联(weakly connected)的。对于无向网络来说,如果每个节点对于其他任何一个节点来说都是可达的,那么这个网络就是连通的,连通度 k 是一旦移走就会出现不再连通的子图或者平凡图(trivial graph)的最少节点数。

包含许多内聚子集和短路径的图通常被称为小世界网络,早期的相关研究来自斯坦利·米尔格拉姆,更近的研究者是邓肯·沃茨。要描述网络中每个行动者中心性的特征,通常可以基于行动者的度(度中心性,degree centrality),或者行动者到所有其他行动者的路径长度(接近度中心性,closeness centrality),或者其他行动者之间的最短路径在多大程度上需要经过给定行动者(介数中心性,betweenness centrality)。网络集中度的测量,表明了在这些不同形式的中心性上行动者之间的异质性程度。

代数技术

与图论方法密切相关的是一些用来理解网络中的社会角色和结构规律的代数技术。对角色进行刻画(characterizations)是用网络映射(mappings)方式进行的,而对结构规律的描述则需要在标记的网络线路(labeled network walks)中进行代数运算。对于"两个行动者具有相同社会角色"的意义有一个重要的命题,蕴含于结构等价的概念之中:如果两个行动者与网络中其他行动者关联的方式以及其他行动者与他们关联的方式都完全一致,那么这两个行动者就是结构等价的(即如果对于所有的 $k \in N$,都有 $x_{ik}=x_{jk}$ 和 $x_{ki}=x_{kj}$,那么节点 i 和 j 是结构等价的)。自同构等价(automorphic equivalence)和规则等价(regular equivalence)等更为一般的形式则基于对 N 的更为一般化的映射,体现出的理念是位置相似的网络节点与相似的其他节点的关联方式相同。

在多个网络中描述结构等价的方法源自以前对亲属体系中结构的刻画方法,可以用多个网络中标记的线路来界定。如果存在一个节点序列 $i=i_1$, i_2,⋯, $i_{h+1}=j$,其中的 i_q 与 i_{q+1} 通过类型为 k_q 的连带联结起来,那么这两个节点 i 和 j 就通过类型为 $k_1k_2\cdots k_h$ 的标记线路(labeled walk)连接起来了(注意序列中的节点可以重复,因此线路是比路

径更为一般化的概念)。每一个连带标记的序列 $k_1k_2\cdots k_h$ 都定义了一个衍生网络,其中的连带表明在网络节点对之间存在着某种特定类型的标记线路。这些衍生网络之间的相等和排序关系会生成各种代数结构,包括半群(semigroups)和局部有阶半群(partially ordered semigroups),描述了在多个网络中线路结构和路径结构中存在的规律性。例如,在含 k 类连带的有向网络中的传递性(transitivity)就是一种结构规律,与之相关的观测情况是,两个节点只有在通过一个 k 类线路相连时才可能也通过类型为 kk 的线路相连。

统计技术

对(有向)图的一种简单的统计模型是伯努利【Bernoulli】分布,其中每个边(edge)或连带都在统计上独立于所有其他连带,连带存在的理论概率为 P_{ij}。这一简单模型不仅假定边具有独立性,而且假定各个连带存在的概率都是相同的;其他模型则可以让概率取决于某些结构参数。这些分布作为统计模型已经用了至少40年了,但是由于独立性假设不太现实,因此用处有限。

网络中的偶体分析结构

社会网络现象的统计模型通过几种主要方式来修正其边独立的初始假定。p_1 模型认识到了在社会网中偶体分析结构(即变量 x_{ij} 和变量 x_{ji} 之间的互依性)在理论和经验上的重要性。伯努利偶体分析分布(Bernoulli dyad distribution)及其赋值的、多元的、时变的(time-dependent)的一般形式,对偶体分析间的互惠和交换关系及其历时发展给出了参数表达方法。模型假设每对偶体分析 (x_{ij}, x_{ji}) 独立于其他偶体分析,可以设定为一般的约束形式如下:

$$P(X = x)$$
$$= \pi_{i<j}\exp[\lambda_{ij} + \theta(\sum_{i<j} x_{ij}) + \rho(\sum_{i<j} x_{ij}x_{ji}) + \alpha_i(\sum_j x_{ij}) + \beta_j(\sum_i x_{ij}x_{ji})] \quad (1)$$

其中,θ 是密度参数,ρ 是一个互惠参数(reciprocity parameter),参数 α_i 和 β_j 反映了不同个体在交往范围和受欢迎程度上的差异,λ_{ij} 保证每个并矢的概率加总为1。这是一个对数线性模型【Log-Linear Model】,很容易拟合。这个模型有多种一般形式,包括假设社会位置和网络连带模式的相互依存的随机块模型(stochastic block models),如 p_2 模型那样的混合模型,以及网络潜在空间模型(latent space models)。

网络的零模型

偶体分析独立(dyadic independence)假设仍然令人生疑。因此,另外一些研究就致力于评估数据对偶体分析间彼此独立这样简单的结构假设的偏离程度与偏离方向。研究引入了一些条件均匀随机图分布(conditional uniform random graph distributions)作为零模型,来探索社会网络的结构特征。这些分布可以用 $U|Q$ 表示,Q 是有向图的状态空间 Ω_n 的一个子集,在 Q 中的每个成员都有相同的概率。子集 Q 通常被选定为有一些特定属性(如一定数量的相互关系、非对称关系、无关系)。如果 Q 等于 Ω_n,那么这个分布就是均匀(有向)图分布,等同于连带概率同质的伯努利分布。列举出 Q 的成员并模拟出 $U|Q$ 是很容易的;但是在某些情况下,例如当分布取决于网络中每个节点 i 的点入度和点出度时,就需要更复杂的方法。

这些分布的一个典型应用就是评估一个网络中特定的高阶(例如三人关系的)特征的出现是否是异常的,前提假定是数据服从基于合理的低阶(如偶体分析的)特征而

建立的均匀分布。已经有更为一般化的测量方法来分析多个网络。最有名的例子可能是弗兰克·贝克和拉里·休伯特的网络二次指派程序（quadratic assignment procedure，QAP）。在这种情况下，对同一组节点而定义的两个网络图的关联可以用均匀多重图分布（uniform multigraph distribution）来估计，而这个均匀多重图分布又取决于每个网络图的未标记的图结构（unlabeled graph structure）。

网络中的超双人局部结构

在社会网参数统计模型的发展中重要的一步是由弗兰克和施特劳斯（Frank and Strauss，1986）作出的，他们引入了马尔科夫随机图（Markov random graphs）。这类模型对于超双人局部结构形式进行参数化，明确地将一些重要的理论观点与统计学网络模型联系起来。这类模型的基础在于，设定在其他随机变量值给定时连带随机变量对 (X_{ij}, X_{k1}) 条件相依，则 Hammersley-Clifford 定理可以给出 $\boldsymbol{X}$ 的一般概率分布形式。

具体而言，有依赖图（dependence graph）$\mathbf{D}$，其节点组为 $N(\mathbf{D}) = \{(X_{ij} : i, j \in N, i \neq j)\}$，边组为 $E(\mathbf{D}) = \{(X_{ij}, X_{k1}) : X_{ij}$与$X_{k1}$在其余 $\boldsymbol{X}$ 给定时条件相依$\}$。弗兰克和施特劳斯用 $\mathbf{D}$ 得到了 $\Pr(\boldsymbol{X}=x)$ 的模型，后来的研究者用 p^* 表示此模型，其中的参数和子结构对应于 $\mathbf{D}$ 的团（clique）。模型形式如下：

$$\begin{aligned}\Pr(X = x) &= p^*(x) \\ &= (1/c)\exp\Big[\sum_{P \supseteq N(D)} \alpha_P z_P(x)\Big] \qquad (2)\end{aligned}$$

其中：

1.需要对于 $\mathbf{D}$ 的所有团 P 进行加总［$\mathbf{D}$ 中的团定义为 $N(\mathbf{D})$ 的一个非空子集 P，使得$|P| = 1$ 或者对于 $X_{ij}, X_{k1} \in P$ 来说有 $(X_{ij}, X_{k1}) \in E(\mathbf{D})$］。

2.$z_P(\boldsymbol{x}) = \prod X_{ij} \in Px_{ij}$是对应于 $\mathbf{D}$ 中的团 P 的（观测）网络统计量。

3. $c = \sum_x \exp\left\{\sum_p \alpha_P z_P(\boldsymbol{x})\right\}$ 是一个标准化的量。

一种相依假定就是马尔科夫假设，即如果$\{i, j\} \cap \{k, l\} \neq \varnothing$，则有 $(X_{ij}, X_{k1}) \in E(\mathbf{D})$。这一假设意味着，从一个节点到另一个节点是否存在网络连带，条件性地取决于这个连带的局部邻域（local neighborhood）中其他连带是否存在。X_{ij}的马尔科夫局部邻域包括所有可能涉及 i 和/或 j 的连带组成。相依假定也可能是其他类型的，在任何建模过程中如何确定哪一种相依假定更为合适，这在理论上仍然是难题。

随机图模型能够对单变量社会网中的局部结构的许多重要概念进行参数化，包括传递性、局部聚类、度分布和集中度。此外，对其进行赋值、多重、加入时序来进行扩展之后，模型也可以对非常有趣的多重关系（multirelational）概念进行参数化，比如与平衡和聚集性（clusterability）、广义传递性（generalized transitivity）和交换、弱连带的力度等相关的概念。伪最大似然估计（pseudo maximum likelihood estimation）较为简单；最大似然估计虽然较为困难，但也不是不可能。

动态模型

一项重要的挑战是提出模型来解释网络现象的出现，包括网络的演化，在关系连带长期存在的背景下个体行动的呈现（如投票、态度转变、决策）和人际交易（如交流模式或者人际互动）。在为网络在离散或连续的时间段中的演化进行建构时，早期的尝试会假定偶体分析之间是独立的且在时间上服从马尔科夫过程。汤姆·思奈德斯及其同事放松了偶体分析间独立的假设，在关于网络演化的连续时间马尔科夫链【Markov

Chain】模型的研究中迈出了重要一步。这种方法也说明了模拟技术能够使得模型假定在经验上更为合理;很显然,这些方法指出了未来发展的前景所在。在网络分析中,基于模拟的计算模型越来越流行;但是,相关的模型评估方法的发展面临着巨大挑战。

卡尔顿、斯科特和瓦塞尔曼(Carrington, Scott, & Wasserman, 2003)对最近的研究(截至2003年)和未来的挑战进行了介绍,包括复杂模型参数的统计估计、模型评估、纵贯数据的动态统计模型等。在此提及的技术和定义的应用实例可以在斯科特(Scott, 1992)和瓦塞尔曼和福斯特(Wasserman & Faust, 1994)的著作中找到。

——Stanley Wasserman

Philippa Pattison

(李洋译　高勇校)

作者注:本研究得到了 U.S. Office of Naval Research 和 Australian Research Council 的支持。

参考文献

Boyd, J. P. (1990). *Social semigroups: A unified theory of scaling and blockmodeling as applied to social networks.* Fairfax, VA: George Mason University Press.

Carrington, P. J., Scott, J., & Wasserman, S. (Eds.). (2003). *Models and methods in social network analysis.* New York: Cambridge University Press.

Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association, 81*, 832-842.

Friedkin, N. (1998). *A structural theory of social influence.* New York: Cambridge University Press.

Harary, F., Norman, D., & Cartwright, D. (1965). *Structural models for directed graphs.* New York: Free Press.

Monge, P., & Contractor, N. (2003). *Theories of communication networks.* New York: Oxford University Press.

Pattison, P. E. (1993). *Algebraic models for social networks.* New York: Cambridge University Press.

Scott, J. (1992). *Social network analysis.* London: Sage.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications.* New York: Cambridge University Press.

Wasserman, S., & Galaskiewicz, J. (Eds.). (1994). *Advances in social network analysis.* Thousand Oaks, CA: Sage.

Watts, D. (1999). *Small worlds: The dynamics of networks between order and randomness.* Princeton, NJ: Princeton University Press.

Wellman, B., & Berkowitz, S. D. (Eds.). (1997). *Social structures: A network approach* (updated ed.). Greenwich, CT: JAI.

神经网络(Neural Network)

神经网络是一种自适应(adaptive)统计模型,可以类比于大脑的结构。说它们是自适应的,是因为它们可以每次使用少量范例(一个或一些)来学习如何估计总体参数【Parameter】。它们与标准的统计模型【Model】并没有本质区别。例如,你会发现神经网络结构与判别分析【Discriminant Analysis】、主成分分析【principal Components Analysis】、Logistic 回归【Logistic Regression】以及其他技术关系密切。事实上,同样的数学工具可以同时用来分析标准统计模型和神经网络。在许多领域中,神经网络都被当作统计工具,包括心理学、统计学、工程学、计量经济学甚至是物理学。神经和认知科学家还将神经网络用作认知过程模型。

基本上,神经网络由有时被称为神经元

(neurons)或者细胞的简单单元建构而成，这也是一种对实物的类比。这些单元通过一组加权连接(weighted connections)而连接起来。它通过连接权重的修正实现自我学习。每个单元编码【Coding】或者对应着我们想要分析的或者想要用作预测变量【Predictor Variable】的模式的某一特征。

这些网络的单元通常在不同的层次上组织在一起。第一层是输入(input)层，最后一层是输出(output)层。中间层(如果有)是隐藏(hidden)层。待分析的信息输入第一层的神经元，进而传送到第二层的神经元以进行进一步的处理。然后将处理的结果传送到下一层，以此类推直到最后一层。每个单元都从其他单元(或者通过某些设备从外部世界)接收信息，对这些信息进行处理，并转而变成输出单元。

网络的目标是学习或者发现输入和输出模式之间的关系，去分析或发现输入模式的结构。在神经网络中，通过修正不同单元的连接权重来实现学习的过程。按照统计术语，这相当于将单元之间的连接值(value of the connections)解释为待估参数(例如，回归【Regression】方程 $y=a+bx$ 中 a 和 b 的值)。学习过程设定了用来进行参数估计的算法【Algorithm】。

神经网络的构件

神经网络由分布在不同层上的基本单元(图1)组成。一个单元收集了由其他单元(或者是外部世界)提供的信息，单元之间通过被称作神经元突触的加权连接相互连接起来。这些被称为突触权重(synaptic weights)的权重与输入信息相乘(或者增加或者减少)：正权重被视为增强性的，负权重是抑制性的。

每个单元都是神经元的简化模型，它将输入信息转换为输出响应。转换包括两个步骤：第一步，计算神经元的激活(activation)值，即输入量的加权总数；第二步，通过转换函数(transfer function)将激活值转变成响应。用公式表达的话，如果输入是 x_i，权重为 w_i，那么激活函数就是 $a=\sum x_i w_i$，同时输出是 o，那么就有 $o=f(a)$。任意函数只要域为实数，都可以成为转换函数。最常见的函数是线性函数($o\propto a$)；阶梯函数(激活值小于给定的阀值，设为0或者-1，其他值设为+1)；Logistic 函数$\left[f(x)=\dfrac{1}{1+\exp\{-x\}}\right]$，取值范围为[-1,+1]，其导数用来进行学习修正，即$\{f'(x)=f(x)[1-f(x)]\}$；以及正态或者 Gaussian 函数$\left[o=(\sigma\sqrt{2\pi})^{-1}\times\exp\left\{-\dfrac{1}{2}\left(\dfrac{a}{\sigma}\right)^2\right\}\right]$。其中一些函数含有概率变异；例如，当激活函数大于给定的阀值时，神经元有1/2的概率可以将其激活函数转换为响应+1。

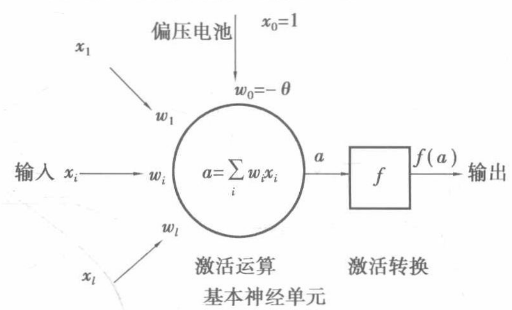


图1 从输入信息到输出信息的基本神经单元过程

网络的架构(即连接模式)与神经元和突触权重采用的传递函数一起，就完全界定了网络的行为。

学习法则

神经网络是自适应的统计手段。这意味着它们像函数一样通过不断改变参数值(也就是突触权重)表现出不同的特征。这些变化依据学习法则(learning rules)做出，这些学习法则可以分为监督(supervised)学

习(如果输出值已知并可以此来计算误差信号)和无监督(unsupervised)学习(如果没有使用误差信号)。

Widrow-Hoff 法则也称为梯度下降(gradient descent)或德塔尔法则(delta rule),是最为人熟知的监督学习法则。它利用神经元实际输入和所要求的输出之间的差异,将其作为输出层单元的误差信号。隐藏层中的单元不能直接计算出误差信号,但是它们能将误差信号估计为下一层单元的一个误差函数(即加权平均数)。这种 Widrow-Hoff 学习法则的自适应也称为误差反向传播算法(error back propagation)。在 Widrow-Hoff 学习法则下,对突触权重(synaptic weights)的修正与误差信号和激活值的乘积存在比例关系,激活值通过传递函数求导得到。引入导数的效应在于,一是当激活接近其极值(最大值或者最小值)时,可以进行细微的修正,二是当激活值处于中间范围时,可以进行较大的修正。如果单元输入相同,那么每种修正都会产生使误差信号变小(smaller)的瞬时效应。一般来说,检视的学习法则在执行最优算法时与下降算法相类似,因为它们都以获得一组自由参数(即突触权重)值为目标,也就实现了整个网络的计算误差函数的最小化。

Hebbian 法则是最为人熟知的无监督学习法则。这个法则是在加拿大神经心理学家唐纳德·赫布相关研究工作的基础上提出来的,他提出神经元学习(即突触变化)是局部现象,可以用神经元激活值之间的时域相关进行描述,并将这一说法理论化。突触的变化同时取决于突触前状态和激活值,因此突触权重的改变是突触前和突触后活动两者间的时域相关函数。具体而言,两个神经元之间突触权重的值在它们处于同一状态时就会增大,在它们处于不同的状态时就会减小。

重要的神经网络结构

在神经网络中最常见的结构是多层感知器(multilayer perceptron)(图 2)。这种结构的大多数网络以 Widrow-Hoff 法则作为其学习算法,并以 Logistic 函数作为隐藏层单元的传递函数(一般来说,这些神经元的传递函数是非线性的)。这些网络之所以常见,是因为它们近似于任何把输入和输出关联起来的多元函数。在统计学的框架中,这些网络与多元非线性回归【Multivariate Nonlinear Regression】类似。在输入模式和输出模式相同的情况下,这些网络就叫作自动关联器(auto-associators)。它们与线性(如果隐藏单元是线性的)或非线性(如果隐藏单元不是线性的)主成分分析【Principal Component Analysis】,以及其他与一般线性模型【General Linear Model】(参见 Abdi, Valentin, Edelman, & O'Toole, 1996)相关联的统计技术如判别分析【Discriminant Analysis】或者对应分析【Correspondence Analysis】关系密切。

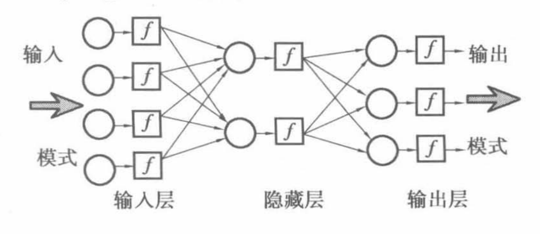


图 2 多层感知器

最近的研究将径向基函数(radial basis function, RBF)网络(见 Abdi, Valentin, & Edelman, 1999)进行了一般化,在支持向量机或者 SVM(Support Vector Machine)(见 Schölkopf & Smola, 2003)的名义下将其整合到了统计学习理论(见 Vapnik, 1999)中。

在这些网络中,隐藏单元被称为支持向量(support vectors),表示可能的(或甚至是真实的)输入模式,隐藏单元的响应是它们与输入模式的相似度的函数。相似度则可以通过核函数(kernel function)进行估计,如点积(dot product);在径向基函数中,核是支持向量和输入量之间欧几里得距离的高斯转换。在 RBF 网络中——我们将其作为 SVM 的例子——隐藏层单元的输出量与由线性单元组成的输出层相连接。事实上,这些网络通过将非线性近似值拆分为两个更简单的值而发挥作用。第一步是简单的非线性映射(即对核到输入模式距离的高斯转换),第二步是从隐藏层到输出层的线性转换。学习主要发生在输出层。这些结构的主要难题在于支持向量的选择以及所使用的特定核。这些网络可以用于模式识别(pattern recognition)、分类和聚类数据。

校验

从统计学的观点来看,神经网络代表了一类非参数自适应模型。在这个框架下,一个重要的问题是对模型效果进行评价。这就要把数据分成两个数据集:训练集和验证集。训练集用来计算网络参数(即突触权重值)。然后,学习过程停止并用验证集数据对网络进行评估。这种交叉验证的方法与自助抽样推论法【Bootstrapping】或刀切法【Jackknife Method】类似。

有价值的参考文献

神经网络理论涉及神经科学、工程学、统计理论等多个领域。不同领域的文献都只是面向某一个读者群体,因此文献来源的多样化也使得研究资料呈现出异质性。对神经网络的统计特征感兴趣的读者可能会发现下面的参考文献有一定价值:Abdi(1999);Bishop(1995);Cherkassky & Mulier(1998);Duda, Hart & Stork(2001);Hastie, Tibshirani & Friedman(2001);Looney(1997);Ripley(1996);Vapnik(1999)。

——Hervé Abdi

(李洋译 高勇校)

参考文献

Abdi, H., Valentin, D., & Edelman, B. (1999). *Neural networks.* Thousand Oaks, CA: Sage.

Abdi, H., Valentin, D., Edelman, B., & O'Toole, A. J. (1996). A Widrow-Hoff learning rule for a generalization of the linear auto-associator. *Journal of Mathematical Psychology*, *40*, 175-182.

Bishop, C. M. (1995). *Neural networks for pattern recognition.* Oxford, UK: Oxford University Press.

Cherkassky, V., & Mulier, F. (1998). *Learning from data.* New York: Wiley.

Duda, R., Hart, P. E., & Stork, D. G. (2001). *Pattern classification.* New York: Wiley.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning.* New York: Springer-Verlag.

Looney, C. G. (1997). *Pattern recognition using neural networks.* Oxford, UK: Oxford University Press.

Ripley, B. D. (1996). *Pattern recognition and neural networks.* Cambridge, UK: Cambridge University Press.

Schölkopf, B., & Smola, A. J. (2003). *Learning with kernel.* Cambridge: MIT Press.

Vapnik, V. N. (1999). *Statistical learning theory.* New York: Wiley.

定类变量(Nominal Variable)

这种变量仅仅是对个案进行分类。它可以让我们说,这些个案属于不同的类别,仅此而已。例如,我们不能像对待定序变量一样,以任何方式对类别进行排序。定类变量的一个例子是选举行为。我们可以根据其在最近一次选举中所投票的政党,对样本进行分类,但是我们不能据此分类对人们之间的差异进行更多的推论。

——Alan Bryman

(李洋译　高勇校)

＊也可参见属性【Attribute】、分类【Categorical】、离散【Discrete】。

通则知识(Nomothetic)

参见通则知识/特殊知识【Nomothetic/Idiographic】。

通则知识/特殊知识(Nomothetic/Idiographic)

普遍规律涉及科学中对一般模型和规律的构建,特殊规律涉及特定情境的具体理解。上述区分是1894年威廉·温德尔班作出的(Smith,1998,pp.141-142)。这种区分是想要划分出自然(通则知识)科学和文化(特殊知识)科学之间的界限,但是界限的位置在哪里,甚至能否如此表述,从那时起到现在一直多有争论。

温德尔班认为尽管所有科学和知识都要面对相同的社会现实,但学科不同因此关注点不同,相应地对社会现实的看法也不同。虽然这不是"意识"和"物质"之间确定性的区分(参见自然主义【Naturalism】),他仍然认为在实践中文化世界和物理世界存在显著差别,文化世界产生于独立的、反身性(self-reflecting)的作为能动者的人。对物理世界的了解需要通过作为观察者的人的客观观察,但是社会世界必须通过解释的方法来进行主观研究。

在20世纪中期到晚期,阐释主义者【Interpretivism】和实证主义者【Positivism】的一场影响深远的争论正是来源于温德尔班的上述区分,但有讽刺意义的是,阐释主义社会科学的一位重要人物——马克斯·韦伯(Weber,1975),为不可僵化地看待普遍规律和特殊知识的区分提出了最佳的论证。他认为这一区分并非是一种科学研究方式和非科学研究方式的区分,他认为两者都是科学的研究方式。韦伯把前者等同于抽象的一般化,如类似法则的陈述。他把后者看作在特定情景下关于具体实际(concrete reality)的科学。

如果持韦伯的观点,那么事实上所有科学,无论是自然科学还是社会科学,都有普遍规律和特殊知识的特征,两者的科学研究方式也能反映出这一点。例如,天文学家能对特定天体单独进行详细的描述,但是对其解释就要将其放在针对广泛现象的抽象一般化的情境中进行。同样,对具体的人类互动的解释需要在更广阔的社会场景中进行情境化(contextualization)。但后者中所需要的某种分类学和理想化,则必须立足于特

殊化的阐释和描述。“20%领取养老金的人生活在贫困中”，这样一种统计表述无法告诉我们任何关于领取养老金个体的信息，而只是抽象的领取养老金的群体，但是“贫困”这一描述却可能是从某一具体的贫困个案的考察中得出的。

自然科学的哲学和社会学也对两者之间的截然区分提出了挑战。温德尔班最初的划分，以及在某种程度上韦伯对其进行的重新定义，都未能触及自然科学的实证主义立场，如价值无涉(value free)和现象主义。但是，从1960年代早期波普尔和库恩的“革命”开始(参见 Lakatos & Musgrave, 1965)，大多数人都认为至少在某种程度上，主观性通过道德定位(moral positioning)、理论选择和观察进入了自然科学。最近几年，理解物理体系和社会体系的复杂方法(参见 Byrne, 1998)强调，这两个体系之间在本体论上具有连续性和相似性，因此反对温德尔班提出的这种认识论和方法论上的区分。

——Malcolm Williams

(李洋译　高勇校)

参考文献

Byrne, D. (1998). *Complexity theory and the social sciences: An introduction*. London: Routledge.

Lakatos, I., & Musgrave, A. (Eds.). (1970). *Criticism and the growth of knowledge*. Cambridge, UK: Cambridge University Press.

Smith, M. (1998). *Social science in question*. London: Sage.

Weber, M. (1975). *Roscher and Knies*. New York: Free Press.

非加性(Nonadditive)

非加性(nonadditive)，这一术语用来描述将一组取值(如变量 Y 的取值)与其他两组及以上取值(变量 X_1 和 X_2 的取值)关联起来的一个函数。如果 Y 可以表示为其他变量的加总，那么 Y 和其他变量之间的关系就是加性的。这个和可以是加权加总，其一般表达式为

$$Y = a + w_1X_1 + w_2X_2 + \cdots + w_kX_k$$

其中 a 和 w 都是常数。

在某些情况下，会对常数项有一些限制条件。例如，如果 $a=0$，权重都等于 1.0，那么变量 Y 就是变量 X 之和。研究者可能会用到的另一种限制条件是权重之和必须等于 1.0，即

$$w_j = \frac{c_j}{\sum c_j}$$

其中 c_j 是变量 X_j 的绝对权重(absolute weight)，加总是对 k 个权数进行的。例如，假设 Y 是 X_1 和 X_2 的加性函数，$a=0$，$c_1=1$，$c_2=1$。那么，$w_1 = 1/(1+1) = 0.50$，$w_2 = 1/(1+1)=0.50$，由此得到

$$Y = 0 + 0.50X_1 + 0.50X_2$$

要注意这个等式与下面等式等价：

$$Y = (X_1 + X_2)/2$$

因此，Y 是 X_1 和 X_2 的平均数。这个例子表明，把 Y 表示为一组变量的平均数是加性函数的一种类型。

非加性函数是 Y 不能表示为其他变量的(加权)和的函数。非加性函数可以有多种形式——实际上，有无穷多的非加性函数形式。非加性函数的一个例子是乘积

【Multiplicative】函数，其中 Y 等于两个变量 X_1 和 X_2 的乘积。

传统的多元回归分析【Multiple Regression Analysis】也是基于加性模型的：

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

其中 α 是截距，β 是回归系数【Regression Coefficients】。

——James J. Jaccard

（李洋译　高勇校）

参考文献

Anderson, N. H. (1981). *Methods of information integration theory*. New York: Academic Press.

非线性动力学(Nonlinear Dynamics)

非线性动力学是指用与时间相关的数学结构模型来进行研究的一系列行为。非线性动力学的代数设定可以是线性的（如 $dy/dt = ay$）或非线性（如 $dy/dt = ay^2$）的，其中参数"a"是一个常数。简单来说，基本思想就是随着时间推移，某一个或多个变量会有变化，而这种变化又不能表示为以时间为水平轴的图形中的一条直线。更准确地说，非线性变化是一种不遵循每段相续且等长的时间内变化幅度亦相同的变化模式。非线性动力学这一术语最经常的用法指的是更复杂的情形，即对有纵贯变化的现象采用非线性数学公式（经常包括由多个变量组成的相依方程组）进行建模，这些公式或直接或间接与时间有关。对非线性动力学如何进行线性和非线性数学设定的全面数学介绍，可以参见 Hirsch & Smale，(1974)。从社会科学的视角来看如何处理同一问题及相关案例，可以参见 Brown（1995a，1995b，1991）。

在物理学和自然科学中，非线性动力学通常采用连续时间微分方程模型（continuous time differential equation model）设定，它是上述学科领域中存在的连续时间变迁过程的结果。但是也有例外情况，尤其是在以代际或季节变迁为研究焦点的生物学中。在这些情况下，会用到差分方程法。在社会科学中，非线性动力学通常使用离散时间的差分方程结构来建模，这主要是由于社会科学数据的收集方式（如周期性的选举、每十年一次的人口普查报告、定期开展的调查数据等）。但同样也有例外，在社会科学中经常会遇到如何用连续时间模型分析离散测量数据这一技术难题，目前这一难题已经找到了解决之道（例如，参见 Brown，1995b）。选用连续或者离散时间方法来为社会现象建模，这一决策在一些情境中会有重要后果，针对这些实质性结果的讨论在布朗（Brown，1995b，pp. 13-30）的著作中有所提及。

非线性动力学通常用如下变化过程来描述：常规的、周期性的、混沌的（chaotic）和突变的（catastrophe）。常规过程始于一个分岔（bifurcation），与之前变化过程有结构性区别的崭新变迁过程自此开始。分岔之后，通常先经历一段正反馈的增长，之后变为负反馈，同时增长过程放缓。非线性动力学的常规过程有一个典型例子，就是 logistic 方程 $dy/dt = ay(k-y)$，当变量 y 逐渐接近均衡值 k 时，y 持续平稳地增长。在邻近均衡值时，增长逐渐趋近零。由于外在于这一常规过程的其他变化过程，常规过程最终变得"成熟"，于是开始下一个分岔，这又开启了一个新的变化过程，新的变化过程可能是前

述四种变化中的任意一种(也可以是一个新的常规过程)。

周期性过程是变量的值在特定区间内重复出现的过程。周期性过程通常是周期性地在某一范围内循环,而常规过程是在某一固定点上达到均衡。混沌过程(参见混沌理论【Chaos Theory】)与周期性过程的区别在于,混沌过程中变量取值并不会精确再现,而且变量值不会重复出现绝非是随机因素的结果。混沌过程在变量取值上通常剧烈变动。因而,混沌过程不会周期性地在某一范围内循环的,尽管变量通常会在一个不稳定的均衡点(被称为"奇异吸引子")的可识别邻近区域中"徘徊"。

突变过程(参见突变理论【Catastrophe Theory】)是经历一个与之前存在的动态过程完全不同的、突然的和剧烈变化的过程。地震这一现象显然就可以用突变过程来建模,因为地震是由于板块位置的逐步变化(即常规过程)最终导致其偏离了先前确立的挤压力量均衡。突变过程本质上是先前处于主导位置的均衡及其基础消失了,新的平衡和基础重新在动态系统中处于主导地位。采用社会科学数据来建立突变模型的案例可以参见 Brown(1995a,1995b)。

——Courtney Brown

(李洋译　高勇校)

参考文献

Brown, C. (1995a). *Chaos and catastrophe theories.* Thousand Oaks, CA: Sage.

Brown, C. (1995b). *Serpents in the sand: Essays on the nonlinear nature of politics and human destiny.* Ann Arbor: University of Michigan Press.

Brown, C. (1991). *Ballots of tumult: A portrait of volatility in American voting.* Ann Arbor: University of Michigan Press.

Hirsch, M. W., & Smale, S. (1974). *Differential equations, dynamical systems, and linear algebra.* New York: Academic Press.

非线性(Nonlinearity)

非线性是指两个或更多变量之间的非线性关系【Relationship】。非线性关系不能用两个维度上的一条直线或三个维度上的一个平面来表示。两个变量 X 和 Y 之间的线性关系可以用一个直线方程表示,即 $Y=a+bX$,其中 a 和 b 分别是截距和斜率系数。但是在社会科学中事物之间的关系很少是决定性的,因此回归【Regression】设定中通常要包含误差项来解释 Y 中未被解释的变差。双变量线性回归【Linear Regression】模型方程,设定了两个变量 Y 和 X_1 的线性关系,其误差项为 ε,截距和斜率参数【Parameter】分别是 β_0 和 β_1,因此等式可以表示为 $Y=\beta_0+\beta_1X_1+\varepsilon$。多元【Multivariate】线性回归模型的一般形式是 $Y=\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k+\varepsilon$,当方程中的其他 X 为常数时,这个方程设定每个 X 和 Y 之间存在线性关系。

回归模型中的非线性是指不能用这些线性模型【Model】来定义的其他关系模式。因而,非线性可以有很多形式。它可以是曲线的,用二次多项式表示为 $Y=\beta_0+\beta_1X_1+\beta_2X_2^2+\varepsilon$,或者是三次多项式表示为 $Y=\beta_0+\beta_1X_1+\beta_2X_2^2+\beta_3X_3^3+\varepsilon$。它可能包含超越函数【Function】,如包括自然对数【Logarithm】(ln)(如 $Y=\beta_0+\beta_1X_1+\beta_2\ln X_2+\varepsilon$),或者自然指数($e$)(如 $Y=\beta_0e^{\beta_1X_1}+\varepsilon$)。它也可能是多元乘积【Multiplicative】的,例如对数线性模型

【Log-Linear Model】$Y=\beta_0 X_1^{\beta_1} X_2^{\beta_2}, \cdots, X_{\beta k}^{k}\ e^{\varepsilon}$。

对"内在线性"的非线性模型和"内在非线性"的非线性模型进行区分很重要。内在线性模型是通过对 Y 和/或 X 进行转换【Transformations】可以简化为线性模型的非线性模型。这相当于用转换后的变量代替初始变量来建构线性回归模型。常见的内在线性模型和线性转换过程参见表 1。

表 1 非线性模型的线性转化

非线性形式	线性转换	线性形式
$Y=\beta_0+\beta_1 X_1^2+\varepsilon$	$X_1'=X_1^2$	$Y=\beta_0+\beta_1 X_1'+\varepsilon$
$Y=\beta_0+\beta_1 \ln X_1+\varepsilon$	$X_1'=\ln X_1$	$Y=\beta_0+\beta_1 X_1'+\varepsilon$
$Y=\beta_0 e^{\beta_1 X_1}\varepsilon$	$Y'=\ln Y$	$Y'=\beta_0'+\beta_1 X_1+\varepsilon'$
	$\beta_0'=\ln \beta_0$	
	$\varepsilon'=\ln \varepsilon$	
$Y=\beta_0 X_1^{\beta_1} X_2^{\beta_2} e^{\varepsilon}$	$Y_1'=\ln Y$	$Y'=\beta_0'+\beta_1 X_1'+\beta_2 X_2'+\varepsilon$
	$\beta_0'=\ln \beta_0$	
	$X_1'=\ln X_1$	
	$X_2'=\ln X_2$	

内在线性的非线性模型相对于变量来说是非线性的,相对于待估参数【Parameter】来说是线性的。相反,内在非线性模型相对于变量和参数来说都是非线性的。在转换完成后,内在线性模型的参数可以用最小二乘【Ordinary Least Squares, OLS】来估计。而内在非线性模型需要更为复杂的技术进行参数估计。一般来说,对这种模型并没有类似 OLS 那种唯一的将误差平方和【Sum of Squared Errors, SSE】最小化的参数估计。相反,要通过一系列不断改进的猜测来重复进行参数估计。

如表 1 所示,非线性模型 $Y=\beta_0 e^{\beta_1 X_1}\varepsilon$ 可以通过双侧取自然对数的方法来线性化,即 $\ln Y'=\ln \beta_0+\beta_1 X_1+\ln \varepsilon$。尽管如此,如果非线性模型假定有一个加数误差项(additive error term)而非乘数误差项,即 $Y=\beta_0 e^{\beta_1 X_1}+\varepsilon$,那么这个模型不能进行线性转换,必须用非线性技术而非 OLS 来估计。另外一个内在非线性模型的例子是含二分【Binary】结果变量的 Logistic 回归【Logistic Regression】模型。

$$Y=\frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\cdots+\beta_k X_k)}}+\varepsilon$$

——Jani S. Little

(李洋译 高勇校)

参考文献

Devore, J. L. (1991). *Probability and statistics for engineering and the sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Greene, W. H. (2000). *Econometric analysis* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Hamilton, L. C. (1992). *Regression with graphics: A second course in applied statistics*. Pacific Grove, CA: Brooks/Cole.

Kmenta, J. (1986). *Elements of econometrics* (2nd ed.). New York: Macmillan.

非参数随机效应模型(Nonparametric Random-Effects Model)

随机效应模型是处理相依观测(dependent observations)的几种方法之一,比如在重复测量和多层次数据结构中出现的相依观测。非参数随机效应模型与标准(参数)随机效应模型的不同之处在于,它并不对随机效应的分布状况进行假设。事实上,这是潜类分析【Latent Class Analysis】的一种形式:混合分布通过有限混合结构

(finite mixture structure)的方式得以模型化。较早的非参数方法文献包括 Laird, 1978; Heckman & Singer, 1982。

令 y_{ij} 表示要研究的响应变量,其中下标 j 表示组别,i 表示组内进行的多个观测(replication)。注意,在重复测量的方法中,组指的是某个个体,多个观测指的是多个时间点。在广义线性模型【Generalized Linear Model】的框架内解释随机效应模型最为简单。假定响应变量服从某个指数族分布,y_{ij} 的期望值 $E(y_{ij})$ 在经过适当的 $g[\cdot]$ 转换后可以用线性函数来模型化。

不含预测变量(predictors)的简单随机截距模型是:

$$g[E(y_{ij})] = \alpha_j$$

如果要使用参数方法,就需要设定 α_j 的分布形式,通常设定为正态分布:$\alpha_j \sim N(\mu, \tau^2)$。需要估计的未知参数是均值 μ 和方差 τ^2。这等于如下参数化方式:$g[E(y_{ij})] = \mu + \tau u_j$,其中 $u_j \sim N(0,1)$。

非参数方法用一个含 C 个节点(即潜在类别)的离散混合分布来描述 α_j 的分布形式,其中的潜在类别用 x 表示,$x = 1, 2, \cdots, C$。与类别 x 相关的截距项表示为 α_x,类别 x 的规模表示为 $P(x)$。α_x 和 $P(x)$ 有时被称为节点 x 的位置(location)和权重(weight)。非参数模型可以设定如下:

$$g[E(y_{ij} \mid x)] = \alpha_x$$

非参数最大似然估计量可以通过增加潜在类别数直到达到饱和点来得到。但是,实践中研究者宁愿使用低于最大类别数的求解。

一旦意识到参数 α_x 和 $P(x)$ 可以用均值(μ)和计算随机效应的方差(τ^2),即参数方法中的未知参数,参数方法和非参数方法之间的相似性就一目了然了。使用初级统计学就可以得到 $\mu = \sum_{x=1}^{C} \alpha_x P(x)$,同时 $\tau^2 = \sum_{x=1}^{C} (\alpha_x - \mu)^2 P(x)$。

如果预测变量是 z_{ijk},就可以得到更一般的包括随机截距的双水平回归模型:

$$g[E(y_{ij} \mid \mathbf{z}_{ij}, x)] = \alpha_x + \sum_{k=1}^{k} \beta_k z_{ijk}$$

其中一个特例是(半参数的)Rasch 模型,它通过为组内的每个观测 i 增加虚拟预测变量而得到。

同时,不同潜在类别的回归系数是可以不同的,这类似于在多层次分析【Multilevel Analysis】中存在随机斜率。由此得到

$$g[E(y_{ij} \mid \mathbf{z}_{ij}, x)] = \alpha_x + \sum_{k=1}^{K} \beta_{kx} z_{ijk}$$

这个模型通常被称为潜在类别(LC)或者混合回归模型。事实上,这是潜在类别分析的一个最重要的应用:发现在回归模型参数上存在差异的不可观察的内部组别。

有两个计算机程序可以估计 LC 回归模型,分别是 GLIMMIX 和 Latent GOLD。

——Jeroen K. Vermunt

Jay Magidson

(李洋译　高勇校)

参考文献

Heckman, J. J., & Singer, B. (1982). Population heterogeneity in demographic models. In K. Land & A. Rogers (Eds.), *Multidimensional mathematical demography*. New York: Academic Press.

Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixture distribution. *Journal of the American Statistical Association*, *73*, 805-811.

Lindsay, B., Clogg, C. C., & Grego, J. (1991). Semiparametric estimation in the Rasch model and related models, including a simple latent class model for item analysis. *Journal of the American*

Statistical Association, *86*,96-107.
Wedel, M., & DeSarbo, W. S. (1994). A review of recent developments in latent class regression models. In R. P. Bagozzi (Ed.),*Advanced methods of marketing research* (pp. 352-388). Cambridge, UK: Basil Blackwell.

非参数回归(Nonparametric Regression)

参见局部回归【Local Regression】。

非参数统计(Nonparametric Statistics)

实验设计、准实验设计【Quasi-Experiment】、事后回溯(expost facto)设计和相关设计经常生成的并不是连续数据(定距【Interval】或比率),而是定类数据(nominal data)或者定序数据。因此,参数统计就不再是假设检验或者描述变量关系的有效方法了。有必要设计一些不再需要参数检验的那些限定的检验和测量方法,其中有些检验并不要求数据是连续性或者服从正态分布【Normal Distribution】(因此有时又被称作"分布无关")。此外,有必要设计一些样本特征是定类的检验方法,此时数据由具有特定特征的样本的频次组成。这些检验自然地被称为非参数检验,它们适于分析定类或者秩/定序数据,目的是以代表性样本【Representative Sample】为基础而进行总体【Populations】推论。

数据及其来源的性质

非参数检验的使用也有其前提要求,而且它的劣势在于与相应的参数检验(parallel parametric test)相比更容易产生第二类错误。非参数检验经常被拿来与相应的参数检验进行功效效率方面的比较。西格尔和卡斯特兰(Siegel & Castellan, 1988)描述了一项功效效率达到90%的检验:"在所有参数检验条件都得到满足的情况下,适当的参数检验即便样本量少10%,也能够达到与非参数检验相同的检验效果。"(p.36)因此,研究者可以令人信服地将数据等级从连续性数据下降到秩数据,用相应的非参数检验取代参数检验,但是检验功效要受到一定影响。表1总结了两种统计类型的需求。

表1 参数和非参数检验的选择标准

参 数	非参数
连续性数据(取值),并且所有组都服从正态分布	定类数据(类别的频次),或者定序数据(等级),或者连续性数据的非正态分布

探讨了参数检验和非参数检验的区别以及非参数检验的限定条件后,我们需要对使用非参数检验的优势予以关注。总结西格尔和卡斯特兰(Siegel & Castellan, 1988)的观点,非参数检验的优势体现在:

(1)适用于潜在分布未知的小样本。

(2)更适用于研究问题。

(3)可以处理秩数据。

(4)可以处理分类(定类)数据。

显著性差异检验

研究设计经常需要确定两组(三组及以上)之间的差异是否比由自然变差引起的差异大。换句话说,无论一组样本如何随机地从总体中抽取,样本并不总是相同。如果它们相差太大,样本就并非来自同一个总体。尽管上述判断在参数检验中很容易理解,可以通过比较不同被试组取值的均值来实现,但如果放在非参数检验中则很难理解,而且两种检验方法缺乏一个共同的主题。在这些检验方法中,感兴趣的差异主要在以下几个方面:

(1)与期望值相比,样本类别的比例(proportions)和频次(例如,男护士和女护士的比例是全国的普遍情况吗?)。

(2)与期望分布相比,取值的分布(distribution)(例如,与中位数相比,不同组的雇员的工资水平)。

(3)离散事件的秩(order)或者顺序(例如,它们真是随机的吗?)。

检验方法的数量和种类是相当多的(见Siegel & Castellan, 1998, 以及Gibbons & Chakraborti, 1992, 后者对检验进行了综述),表2显示了与参数检验相比,相应的非参数检验的用途。如果研究者需要对样本能否代表来自同一总体的所有样本进行检验,那么对于数据(秩)来说,与z检验相似的是Kolmogorov-Smirnov检验【Kolmogorov-Smirnov Test】,对于定类数据(类别的频次)来说,就有卡方拟合优度检验。在这两个检验中,都要对样本和总体进行比较。同样,与参数检验中的t检验相似,在非参数检验中也可以检验两个样本是否来自同一总体。对于定序数据来说,Mann-Whitney U检验【Mann-Whitney U test】可以比较两个独立样本,而对于定类数据来说,卡方检验可以解决两个样本的假设检验问题。对于相关样本或配对样本来说,与t检验相对应的处理相关样本的检验技术是Wilcoxon符号秩检验,如果是定类数据,可以采用McNemar卡方检验【McNemar's Chi-Square Test】(有时候也叫作McNemar变化检验)。如表3所示,对三组及以上组样本,也有与参数检验类似的检验方法。

表2 典型的参数和非参数显著性检验的比较

	自变量或者样本组				
		两组样本		三组样本	
因变量测量	单样本	独立	相关/匹配	独立	相关/匹配
定距/定比	z检验	t检验	$t_{相关}$检验	单维ANOVA	随机块ANOVA
定序	t检验($n<30$) Kolmogorov-Smirnov检验	Mann-Whitney U检验	Wilcoxon符号秩检验	因子ANOVA Kruskal-Wallis单因素方差分析	ANCOVA Friedman秩数据双因素方差分析
定类	χ^2检验,拟合优度检验	χ^2检验,$k\times 2$表格	McNemar变化检验	$m\times k$表格 χ^2检验	Cochran Q检验

表 3 相关和关联的几种典型的参数和非参数测量

			定类	
	定距/定比	定序	$m \geq 2$	$m=2$
定距/定比	皮尔逊积距相关，r_{xy}（线性关系）。如果非线性、单调：两者都简化为定序层次 ↘	（简化为定序层次的定距/定比变量）↓		点双列相关，r_{pb}
定序		斯皮尔曼 rho，r_s（线性关系）	简化为定类层次的定序变量	
定类			↓	
$k \geq 2$			Cramer $C(V, \phi_c)$	
$k=2$				Phi，ϕ（二分变量）

相关和关联的描述量

对单一组的调查结果通常试图描述变量对之间相关（连续性或定序数据）或者关联（定类）关系。统计量的选择取决于数据的性质，以及在这种情况下变量是如何测量的，如表 3 所示。如果两者都相同（例如，都是连续性的），那么结果很简单：就是采用皮尔逊积差相关。但是如果两者并不相同，那么选择的统计量就取决于最低层次的数据。例如，如果一个变量是年龄（连续性），第二个变量是社会阶层（定序），那么就必须也将年龄数据视为定序的（年龄转化为秩），这样，斯皮尔曼 rho 检验是最合适的检验技术（表 3）。

参数和非参数检验比较的一个例子

尽管数据类型是选择参数检验还是非参数检验的决定因素，但对非参数检验所能够适用的研究问题类型、两种检验方法之间的差别，以及参数检验所能解决的问题进行充分考虑还是有必要的。有时候，由于我们能够获取的数据层次有限，因此不得不改变研究问题。为了更好地阐释这一问题，这里将以实例来说明参数检验和非参数检验的区别。

在这个例子中我们将使用两种常见的检验方法来说明其基本差别：t 检验和卡方检验或者 χ^2 检验。后者主要用于定类数据，但是由于其使用简单方便，有时候用于含秩类别频次的定序数据，尽管这可能会轻微增加产生类型Ⅱ误差的概率。

假设这样一个研究情景，研究者对两个样本组就政治观点的保守性进行提问，看两组之间是否存在差别。如果此项研究就政治观点的保守强度进行问卷测量，并得到每个受访者的分值，那么就可以使用 t 检验来比较两个样本组的平均取值是否有差别。另一方面，如果研究者会对每个样本在近期选举所支持的政党进行记录，假设有一个政党是保守的，那么每组选民投票给每个政党的次数就构成了全部数据。为了两组样本在投票模式上的差异，χ^2 检验是一种较为合适的选择。但是要注意，这里回答的其实是两个不同的问题：

- 两组样本在投票模式上有区别吗?
- 两组样本在保守主义程度上有区别吗?

因此,研究者必须意识到,其所收集数据的类型可能会影响要回答的问题。在这两种情况下,人们可能希望推论出两组之间的保守主义是否有差异,但是只有前一项研究能真正回答这一问题。有关这个问题,我们回头会看到一个具体数字的例子,但是首先让我们看下与 z 检验和 t 检验并行的几个非参数检验的例子。

用哪一个:t 检验或者卡方检验

在本词条开始,我们就提出了这个问题。采用何种检验方法的最重要的标准是为回答此研究问题而收集的数据类型。例如,如果问题是"两个英式酒吧的顾客的保守程度有差别吗?"接下来我们就要看需要用什么样的数据来回答这个问题。简单来说,如果样本从全体顾客中随机抽样,且数据是样本的分值,即满足表 1 左侧一列的标准,那么使用参数 t 检验较为合适。但是如果根据顾客保守程度的高低,将样本分成几个类别,然后计算每个类别出现的频次,那么就需要用非参数卡方检验。

具体来说,针对上面的问题我们可能会根据两种对变量进行操作性界定的方法给出两种解答。首先,我们可以设计一种特制的检验方法,根据保守主义水平设定相关的分值,以此来对"多么保守"进行界定。或者说,我们可以本着"行胜于言"的原则,简单询问受访者在选举中的投票政党(或者想要投票的政党),这里,三个主要的英国政治党派的政治理念是从保守主义到社会主义逐渐过渡的(至少在过去是这样的)。如表 4 所示,上面两种设计思路会产生两种不同的检验,从逻辑上来说,它们来自两种不同的统计检验。

表 4 这两种方式看似回答的是同一个问题:"两个酒吧的顾客的保守程度有差别吗?"

参 数	非参数
这个酒吧的顾客比另一个酒吧的更保守吗	两个酒吧的顾客在投票中的党派立场是否相同
用问卷(20 个问题,5 级量表)来测量"保守程度";得分越高,表明越保守;得分的区间在 20~100	在地区选举中,为每个党派投了(或者想投)支持票的人有多少(每个党派得票数的频数表)
用 t 检验来比较两个酒吧的得分均值	用卡分检验来比较两个酒吧的投票频数表

我们甚至可以用假设数据来展示每种方法可能会得到何种结果(如果是定序数据,我们可以使用更敏锐的检验,但结果基本类似)。表 5 中显示了这一点,每种方法都使用了样本数据。这就提出一个关于"保守主义"变量的测量的问题:它们真的是用两种不同的测量工具测量了一个相同的东西吗?或者说,是否存在其他的因素会影响某种测量工具的结果,会歪曲受访者的响应?谁在提问?这是争论的起点。

使用问卷对保守主义进行测量的研究者可能认为测量与谁是竞选者无关,问题在于无论选民的政治倾向如何,竞选者的个性都可能会影响到选民的投票行为,因此对政党的选择并不能有效反映该选民的保守程度。尽管人们一般认为保守党人最保守,自由民主党人居中,劳动党选民最不保守,但在对于具体议题和具体候选人上,上述区分既不合理也难以进行严格排序。但是,询问人们会投票给谁的研究者认为,这种方法可以避免抽象的提问,观察人们对信念的践行,关注到了政党和意识形态。可以肯定的是,检验方法是适合于所收集到的数据的,但还存在一个问题,即哪种方法最适用于回

答研究问题。换句话说,最大的异议在于对操作化定义【Operational Definition】(即测量工具)的选择而非对统计检验的选择,而操作化定义本身又是研究问题的逻辑延伸。

表 5 采用两种方法表面上对相同问题"两个英式酒吧的顾客的保守程度有差别吗?"进行回答的数据和结果

参数			非参数		
	绿蟾蜍	红鲱鱼		绿蟾蜍	红鲱鱼
均值	78	68	保守党	35	20
标准差	22	24	自由共和党	33	36
数量	87	76	劳动党	19	20
			总　计	87	76
$t=2.76(p<0.05)$ 拒绝 H_0:不同保守主义水平上可能有差异			$\chi^2=3.52$(n.s.) 接受 H_0:在选举模式上可能没有差别		

历史

这里根据不同检验方法的开发者和公开日期,对非参数检验发展的简要过程进行重点回顾(参考 Howell, 2002; Kerlinger & Lee, 2000; Siegel & Castellan, 1988)。从卡尔·皮尔逊开始,他在 1900 年左右提出了卡方检验;卡方检验有时也被称为皮尔逊卡方检验。而经济学家的米尔顿·弗里德曼,在 1937 年提出了相关样本秩数据的双因素方差分析。在 1941 年左右,柯尔莫哥洛夫提出了针对定序数据的单样本和两个独立样本比较的拟合优度检验,斯米尔诺夫公开了这种检验的相关表格。威尔科克森(Wilcoxon, 1945),以及曼和惠特尼(Mann & Whiteny, 1947)也提出了其他等效的定序数据的双独立样本检验。1952 年,克鲁斯卡尔和瓦利斯提出了定序数据的单因素方差分析检验。1969 年,奎因·麦克尼马尔概述了针对两个相关定类数据样本的符号检验方法。

——Thomas R. Black

(李洋译　高勇校)

参考文献

Black, T. R. (1999). *Doing quantitative research in the social sciences: An integrated approach to research design, measurement, and statistics.* London: Sage.

Bradley, J. V. (1985). *Distribution-free statistical tests.* Englewood Cliffs, NJ: Prentice Hall.

Gibbons, J. D., & Chakraborti, S. (1992). *Nonparametric statistical inference* (3rd ed.). New York: Marcel Dekker.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Duxbury.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research.* Ft. Worth, TX: Harcourt.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.

非参与观察(Nonparticipant Observation)

以观察为基础的,但不让研究对象意识到正在被研究的研究,被称为非参与研究,或非介入研究,或无反馈研究(nonreactive research)。伪装观察(disguised observation)和自然实验【Natural Experiment】,这两种非参与观察的形式无须研究对象同意,曾经一度为人所接受,但现在被认为在伦理上不可行,因为它们违反了知情同意【Informed Consent】的原则。但是其他非介入性的方法仍然在被社会学家、文化人类学家和社会心理学家使用。

行为追踪研究(behavior trace studies)、档案研究【Archival Research】和内容分析【Content Analysis】是三种常见的非介入性方法,这些方法并不要求研究者连续对实际行动进行在场观察(on-site observation)。采用非介入方法【Unobtrusive Methods】进行在场观察的研究人员会系统性地记录资料,常常会使用预先编码的清单表(checklists)。由于研究人员无法就其行为询问被观察者来确证自己的解释,因此他们必须用最客观的方式,格外仔细地对资料进行记录。同时,他们必须用汇总形式公开其研究结果,以防因泄露个人身份而无意间损害个人的隐私。如果这样的观察是由不同的研究者在不同的地点实施的,还要确保研究者采用一致的格式来记录数据。

有些问题如果直接问,参与者是不会回答的(如像青少年的调情方式这样的隐秘行为,或者像吸毒这样从法律或伦理上来说研究者都不能参与的行为),这时非参与观察就特别重要了。非参与观察已经广泛应用到民族志【Ethnographic】研究中,如描述非言语沟通的特征(即理解人们如何使用空间安排或者肢体动作),或者描述人们作出购买选择的方式(但是并不分析选择背后的原因)。例如,普赖斯(Price,1989)对位于厄瓜多尔的两个城市药店的消费者行为进行了研究。她想要知道是否有些处方麻醉药是在没有处方的情况下被买走的,在观察了超过600个交易业务后,她断定其中51%是在无处方的情况下购买了处方麻醉药。她同时发现在这些例子中,消费者通常只购买少量的胶囊(每次两片或者三片)。此外,普赖斯意外发现药剂师(甚至还包括未受培训的药店职员)是提供用药建议的主要人物,尤其对通常难以见到内科医生的那些低阶层的消费者。

理查德·李(Lee,1993)的一项研究被广泛引用,这项研究关注南部非洲昆申人的食物来源和饮食构成。他着手测量当地人从营地步行到聚会点或狩猎点的距离。同时,他计算或推测出每个劳力每小时采集的坚果或者猎杀的动物的重量、从植物或动物资源上获取的平均卡路里数、消耗的不同食物的数量以及分享不同种类食物的人之间的社会关系。在这些观察的基础上,李能为下一个阶段的研究构想出有关食物、饮食和社会网络的更有意义的问题,这比他仅仅抱有对当地人行为的外在成见而进入田野调查要好得多。

尽管非参与观察者不愿意直接与观察对象进行互动,但有时候通过提问题的方式来得到一些情景信息还是很有必要的;这样的背景材料可以在实际观察开始之前进行收集。同样的,一些进行观察的研究人员发现,有时有必要提问来澄清观察到的行为;这些问题也可以在观察结束之后提出来。但是在非参与观察中,主要的研究进程还是应当在研究人员无干预的情况下进行。

——Michael V. Angrosino

(李洋译　高勇校)

参考文献

Lee, R. B. (1993). *The Dobe ! Kung*. New York: Holt, Rinehart and Winston.

Price, L. J. (1989). In the shadow of biomedicine: Selfmedication in two Ecuadorian pharmacies. *Social Science and Medicine*, *28*, 905-915.

Sechrest, L. (Ed.). (1979). *Unobtrusive measures today*. San Francisco: Jossey-Bass.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences*. Chicago: Rand McNally.

非概率抽样(Nonprobability Sampling)

非概率样本就是未用随机方法选择出来的样本。

抽样的目的是通过从总体中抽取样本【Sample】而得到关于总体的准确代表,以便基于样本对总体进行准确的推论。最具代表性的样本是随机抽取的大样本(参见随机抽样【Random Sampling】),即总体中的每个个体被选中的概率都是已知的且非零的(在简单随机抽样中,每个个体被抽中的概率都是相同的,但也有其他变化形式)。

随机抽样通常是不可能实现的,或者成本很高。它需要调查总体的全部清单,或者采用其他方法来保证每个个体都被纳入抽样范围,即使被选中的单元无法找到或者拒绝参与时也不能对其替换。在这些情况下,调查组织经常采用配额抽样【Quota Sampling】这样一种非概率方法。这样,总体的代表性就不再是用概率决定而是用样本模式决定。例如,如果认为性别变量很重要,相关总体就被分成男性和女性,调查设计要求访谈员选择同样多的男性和女性——他们会有一个性别配额来寻找被访者。如果正确实施,此类样本在配额变量上必定能代表总体。这是由研究设计决定的。

但是,配额抽样在其他方面并不具有代表性,因为访谈员可以自由选择如何完成配额。例如,英国开放大学的学生进行了一项调查作为他们的课程作业,每个学生负责用配额方法收集四个个案,要考虑性别、年龄和社会阶层,结果在全国范围内抽取到了约2 000个样本。当着手分析数据时,我们发现在总体大多数的已知事实上,样本都具有代表性。中产阶层群体不分性别都倾向于投票给保守党人,这是一个全国趋势,但是在此样本中,中产阶层女性却更愿意投票给工党。对此的解释是,学生们在找中产阶层女性时,会就便去自己孩子的学校里找老师,而研究表明,在公共部门工作的中产阶层群体与其他群体更倾向于投票给工党。

配比样本至少能在某些方面代表总体,但是其他形式的非概率样本连这一点也做不到(见方便样本【Convenience Sample】)。例如,许多研究是基于"偶遇(haphazard)"抽样的,如"随机"地在街道上拦住路人。还有一些研究基于"机会样本(samples of opportunity)",如恰巧在你任教的班级进行教育研究。还有一些研究使用自愿样本,做广告来招募参与者。在所有这些情况下,很有可能总体不具有充分的代表性。我们无法遇到在那个时间不在街上的那些人。我们的班级可能无法代表整个学校,更无法代表城镇中的所有人(也无法代表整个国家)。自愿抽样无法接触那些不愿参与研究的人,志愿参与者最有可能是那些对所宣传的事态度坚定的人,那些对此漠不关心的人参与的可能性更低。

因此,要谨慎对待基于非概率样本而来的信息。或者说,研究者需要仔细考虑这些

信息能代表什么总体。例如,有一项关于孤独的调查是通过在咖啡厅做广告得到的自愿样本进行的,它就无法代表那些从不外出去咖啡厅的人,这些人才是真正的社会隔离者,它也无法代表那些由于害羞而无视广告宣传的人。虽然它有上述局限性,但仍有可能提供一些有用的信息。

有一种观点认为,无论设计多么严格,所有的调查获取到的样本都是非随机的。抽样方式可能是随机的,但是无法找到或拒绝参与者不太可能是所计划样本的随机子集;它们和回答者存在着系统性的差异。

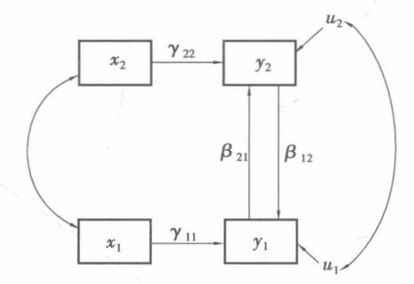
——Roger Sapsford

(李洋译　高勇校)

非递归(Nonrecursive)

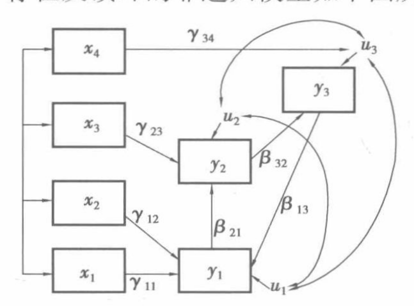
许多社会科学理论的模型是由一个以上方程设定组成。方程的非递归系统是联立方程【Simultaneous Equation】系统的一种类型,在这种类型中,方程通过因变量的双向关系【Reciprocal Relationship】或者方程误差项之间的相关性联系起来。在非递归模型中,或者(a)存在互为因果(reciprocal causation)因素,或者存在反馈环(feedback loops),并且/或者(b)方程中的误差项有相关关系。存在互为因果关系或者误差项相关(或者同时存在)表明模型是非递归的。相反,如果既没有相互因果关系,误差项也不相关,这个联立方程系统就是递归【Recursive】的。相依回归【Seemingly Unrelated Regression】模型是非递归模型的一个特例,其中方程之间的关系只能通过误差项相关体现出来。

一个简单的、存在互为因果关系的非递归模型可以用下面的路径图【Path Diagram】表示:



这里,y_1 影响 y_2,而 y_2 反过来又影响 y_1。因此,y_1 的变化会反馈回自身。同时要注意的是,方程中的误差项是相关的。

存在反馈环的非递归模型如下图所示:



可以看到,y_1 影响 y_2,y_2 影响 y_3,同时 y_3 又反过来影响 y_1。y_1 的变化会反馈影响到其自身。

下面看一下与第一个图示相对应的含有两个方程的非递归方程体系:

$$y_1 = \beta_{12}y_2 + \gamma_{11}x_1 + u_1$$
$$y_2 = \beta_{21}y_1 + \gamma_{22}x_2 + u_2$$

其中,$\mathrm{Cov}(u_1,u_2) \neq 0$。

在方程体系中,内生变量【Endogenous Variable】用 y 表示,外生变量【Exogenous Variable】用 x 表示。β 表示一个内生变量对另一个内生变量的影响,而 γ 表示一个外生变量对一个内生变量的影响。u 表示方程中的误差项。为了实现方程均衡,$|\beta_{12}\beta_{21}|<1$。

方程的非递归方程系统必须要避免识别问题【Identification Problem】。也就是,必

须要确定模型参数是否可估计。可以采用许多方法来识别非递归模型,包括顺序条件和秩条件(参见 Bollen,1989,pp.98-103)。里格登(Rigdon,1995)提出了一种能直接识别多种类型非递归模型的方法。

对方程的非递归方程系统进行估计时,需要考虑系统中其他方程的信息。因而,如果用最小二乘【Ordinary Least Squares,OLS】回归来估计非递归联立方程组,估计可能有偏且不一致。故而,非递归模型应当通过两阶段最小二乘法、三阶段最小二乘法和最大似然法【Maximum Likelihood Estimation】进行估计。这些估计量其实都可以看成某种工具变量【Instrumental Variable】估计量。

——Pamela Paxton

(李洋译 高勇校)

参考文献

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Greene, W. H. (1997). *Econometric analysis* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Rigdon, E. E. (1995). A necessary and sufficient identification rule for structural models estimated in practice. *Multivariate Behavioral Research*, *30*, 359-383.

Heise, D. R. (1975). *Causal analysis*. New York: Wiley.

无回答(Nonresponse)

在调查研究中,通常有两种类型的无回答:单元无回答和题项无回答,前者指被访者没有完成调查问卷,后者指被访者没有回答其中一个或多个问题。

单元无回答

总体而言,对家庭户和企业的调查中,应答率在最近几十年中一直持续下降。无回答增加了数据收集的成本,也可能导致无回答偏倚【Bias】。

格罗夫斯和库珀(Groves & Couper,1998)区分了影响被访者调查合作的内在因素(exogenous factors)和外在因素(endogenous factors)。外在因素多数是调查研究者所无法控制的,包括社会环境(如调查区域的城市化水平或事先对本调查的知晓度)、样本单元的社会心理学特性以及对所涉及议题的了解程度。内在因素包括数据收集规范(填写模式、激励、问卷长度等),在访员执行的调查中还会包括访员的筛选和训练。

在企业调查中赢得合作是很困难的。在很多情况下,没有一个人能够掌握回答问卷所需要的所有信息。在电话研究中,访谈员很难跳过"守门人"直接与被调查者谈话。一些公司对于员工参与调查一概禁止。即便企业没有这种政策,受邀参加调查的人也必须首先得到组织内其他人的授权才能参加调查。

有一些研究设计方式已经被证明对提高回答率有积极作用。与调查样本成员联系的次数和方式是赢得参与者合作的重要因素。邮寄调查通常会寄送提醒明信片、邮寄备用问卷包和给未回答者打提醒电话。在电话调查【Telephone Survey】中,打电话的时机对提高找到在家的潜在参与者的概率很重要。计算机辅助数据收集【Computer-Assisted Data Collection】能综合采用各种规则,确保在白天和晚间、工作日和周末都尝试打电话,同时确保隔周拨打电话。影响回答率的其他设计因素还包括实地田野期的长度(通常越长越好);调查资助人;事先发送关于相关研究信息的信件并让参与者知晓访谈员将要打电话;以

及货币和非货币的激励。先期支付的现金奖励已经证明是最有效的激励类型。对拒访的追访和增加电话尝试的次数也能增加回答率。

无回答者的相关信息通常是有限的。在无回答研究中,会在无回答者中随机抽取出一个样本,然后施以大量努力来吸引其参加(例如提供或增加激励,提供多种不同的填答方式,以及/或者增加联络的次数)。内容通常包括调查中的关键问题和人口特征方面的问题。收集到的数据可以用来概括无回答者的特征,并对数据进行事后调整。

当回答者与无回答者存在系统性差异,且这种差异与调查中的研究问题相关时,无回答偏倚【Nonresponse Bias】就产生了。人们普遍认为无回答是调查研究中误差【Error】的一个主要来源。提高回答率本身是值得重视的,因为它可以提高数据的可信性。但是,提高回答率最重要的好处是有助于减少无回答偏倚。付出额外努力来减少无回答偏倚的效果,取决于回答者和无回答者之间的差异有多大。减少无回答的实质意义因研究主题、总体、数据收集方法的不同而不同。

题项无回答

回答者可能无法准确地遵循跳答模式,会故意忽略使其感到不舒服的问题,或者在问卷完成之前停止答题。这种题项无回答在自填式(self-administered)调查中会更突出,而在有访谈员在场的调查中会少一些。访谈员可以与被调查者建立起融洽的关系,帮助他们克服填写问卷过程中的困难,能够重复问题或者在必要的时候对不熟悉的词汇进行标准化界定。

与威胁性较低的题项相比,不回答更可能出现在敏感问题中。当提到关于收入水平的问题,以及社会不认可的行为,比如为消遣而服用麻醉剂或者犯罪行为时,经常会有缺失数据【Missing Data】。在一项给定调查中,询问被调查者收入情况的题项经常会产生最多的题器无回答。用一系列的嵌套问题来询问收入情况,以及/或者在回答选项中提供收入范围,有助于改善数据质量。

设计较差的问题(如含糊的或者有歧义的题项)经常产生缺失数据,单元无回答和题项无回答都可能出现。被调查者回答很费劲的题项,无回答的概率也会很高。例如,让被调查者回答百分比且总计必须是100%的问题、回忆很久之前情况的题项、自填问卷中的开放式题项。在问卷设计阶段,对那些能与调查总体的关键特征相似的人就备选题项进行认知访谈来测试问卷,进行实地预调查,都将会有助于发现有问题的题项。在调查实施之前,可以修订这些题项。

问卷中的跳答指示(skip instructions)告诉被调查者根据触发题项(trigger item)的回答情况来决定接下来应该回答哪个问题。当被调查者没有回答他们应该回答的问题,或者当他们回答了理应跳答的问题时,就出现错误的跳答。设计周详的问卷能使跳答误差的可能最小。调查研究人员要认真考虑提供给被调查者的有关跳答的图像内容和文本内容。

在对题项无回答进行考察时,应当认识到两种“纯粹”的缺失回答之间的差异,一种是被调查者对此问题尚未形成意见,另一种是这个题项不适用于被调查者。最好使用筛选性问题。这些“触发”问题可以用来确定某一实质性问题的适宜分母是多少,那些缺乏相关经验而无法回答事实题项(factual items)的被调查者,或者那些在意见题器上没有意见的被调查者,可以跳过“目标”题项。但是也有人会担心,提供“无

意见”这一选项可能会给被调查者提供一种逃避回答问题的简便办法,而实际上他们可能有确切的意见。

在企业调查中,当要求被调查者提供公司没有采集的信息时,或者公司采集的信息单位与问卷所提供的回答选项不符时,就会产生题项无回答。同时,企业也会考虑信息的所有权问题——雇员可能被禁止披露某些信息。

能够使回答率最大化的研究设计和能够使问题设计最优化的问卷研制,是减少调查无回答的关键。

——Patricia M. Gallagher

(李洋译 高勇校)

参考文献

Fowler, F. J., Jr. (1995). *Improving survey questions: Design and evaluation*. Thousand Oaks, CA: Sage.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York: Wiley.

Groves, R. M., Dillman, D. A., Eltinge, J. L., & Little, R. J. A. (Eds.). (2002). *Survey nonresponse*. New York: Wiley.

Schwarz, N., & Sudman, S. (Eds.). (1996). *Answering questions: Methodology for determining cognitive and communicative processes in survey research*. San Francisco: Jossey-Bass.

无回答偏倚(Nonresponse Bias)

当一项研究没有收集到从抽样框中所选样本的所有数据,也就意味着出现了无回答【Nonresponse】。无回答偏倚可能会出现在任何形式的研究中,而不仅仅是在抽样调查中。回答率是对无回答的规模进行的测量,回答率的计算有多种方法(参见AAPOR,2000)。尽管回答率能够说明研究是否有效地测量了所有抽中的样本,但是它并不能直接说明这些数据是否会产生无回答偏倚。也就是说,“低”回答率并不意味着就存在无回答偏倚(参见Curtin, Presser, & Singer, 2000; Merkle & Edelman, 2002)。但是,如果在我们要进行的测量上,无回答样本与回答样本有所差别,那么无回答误差就会出现,通常以偏差(bias)的形式出现。无回答偏倚同时取决于以下两个因素:无回答的规模,回答者与无回答者之间的差异大小(Groves, 1989)。

不管回答率是多少,如果回答者和不回答者在要进行的测量上没有差异,那么就不存在无回答偏倚。无回答偏倚可能在单元的层次上出现(如个人),也可能在题项层次上出现(如调查问题)。在调查中,无回答大多是由于无法联系、拒绝和语言障碍等原因而产生的,已经设计出许多方法上的技术来尽量降低无回答的概率。如,许多统计加权【Weighting】和插补法【Imputation】技术可以用来对单元和题项层面出现的无回答进行调整。尽管如此,一般从无回答的样本中收集不到任何信息;因此,我们并不保证上述统计技术能降低可能存在的无回答偏倚。针对无回答者的特定的跟踪研究能收集到一些信息,以便确定最初研究中的无回答者与回答者有怎样的区别。

——Paul J. Lavrakas

(李洋译 高勇校)

参考文献

American Association for Public Opinion Research (AAPOR). (2000). *Standard definitions: Final dispositions for case codes and outcome rates for*

surveys. Ann Arbor, MI: Author.

Curtin, R., Presser, S., & Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment. *Public Opinion Quarterly*, *64*, 413-428.

Groves, R. M. (1989). *Survey errors and survey costs.* New York: Wiley.

Merkle, D. M., & Edelman, M. (2002). Nonresponse in exit polls: A comprehensive analysis. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp.243-258). New York: Wiley.

非抽样误差(Nonsampling Error)

在社会研究中有各种形式的潜在误差,面对这些误差,研究者(a)努力将其最小化并且/或者(b)对其进行测量(参见 Groves, 1989)。无论是定量研究【Quantitative Research】还是定性研究【Qualitative Research】,误差的产生都是由于收集资料时出现的各种原因引起的偏倚【Bias】(即稳定误差)和方差【Variance】(即变异误差)。抽样误差【Sampling Error】是与任何非普查的抽样相关联的方差。非抽样误差传统上是指所有其他的偏差和方差来源,包括涵盖误差、无回答误差以及测量误差。

涵盖误差(coverage error)是指代表总体的抽样框(列表)没有全部覆盖(包括)整个总体而产生的偏差。无回答【Nonresponse】误差是指没有从抽样框所选样本中采集到所有数据而产生的偏差或者方差。测量误差(measurement error)是指由以下几点引起的潜在的偏差或方差:(a)收集数据的人(例如访谈员、观察员或者编码员);(b)用来收集数据的工具(例如问卷、观察或编码表);(c)接受测量的被试/受访者;(d)数据收集模式(例如,面对面访谈问卷还是自填问卷)。目前有很多方法研究和统计研究的文献,介绍了降低非抽样误差发生概率的各类技术,以及/或者当误差出现时对其进行测量和调整的各种方法。要注意的是,很多人认为这些形式的误差只出现在抽样调查中,而实际上每种类型的误差在任何社会研究形式中都有相对应的部分,包括观察研究【Observational Research】、内容分析【Content Analysis】以及各种形式的定性研究。

——Paul J. Lavrakas

(李洋译　高勇校)

参考文献

Groves, R. M. (1989). *Survey errors and survey costs.* New York: Wiley.

正态分布(Normal Distribution)

正态分布(也称为高斯分布)通常用于解决社会科学中的各种问题。许多连续随机变量可以建模为正态概率【Probability】分布,其密度函数生成一个对称的、钟形的曲线。连续随机变量 Y 的正态密度函数是

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

其中，$-\infty < y < \infty$，$-\infty < \mu < \infty$，$\sigma > 0$，且 μ 和 σ 分别是总体【Population】均值和标准差【Standard Deviation】，因此，$E(Y)=\mu$，$V(Y)=\sigma^2$。可以看出，定义正态分布密度只需要这两个参数【Parameter】。

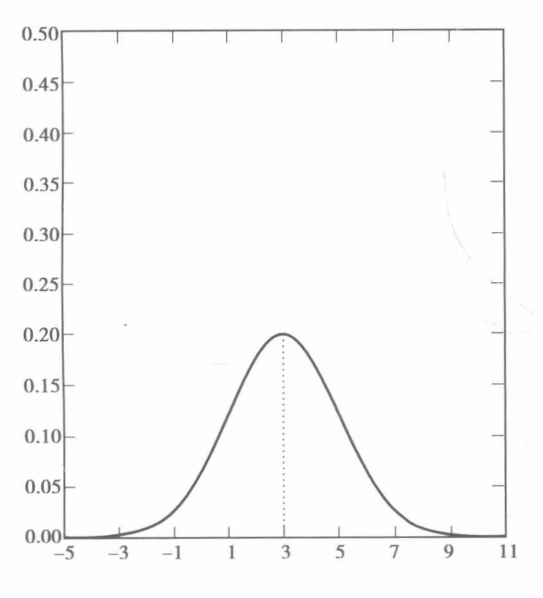


图 1 Y_i 表示正态，$\mu=3$，$\sigma=2$

正态分布有一个特殊的属性，即在不同的参数值下都保持：68%的分布位于均值上下的 1 个标准差范围内，95%的分布位于均值上下的 2 个标准差范围内，99.7%的分布位于均值上下的 3 个标准差范围内。尽管如此，由于分布的取值范围是 $-\infty \sim +\infty$，因此正态分布随机变量在实数线（real-number line）的任意位置上都有非零概率取值。此外，由于分布是对称且有唯一的最大值，因此它的均值【Mean】、中位数【Median】和众数【Mode】都是相同的。再者，如图 1 和图 2 所示，两个反曲点（inflection points）位于离均值 1 个标准差的位置。要注意这两个分布都以 3 为中心，但是第一个分布的延展要大于第二个分布。尽管如此，在两个分布中，都有可能（即使是很低的概率）出现很高或很低的数字。

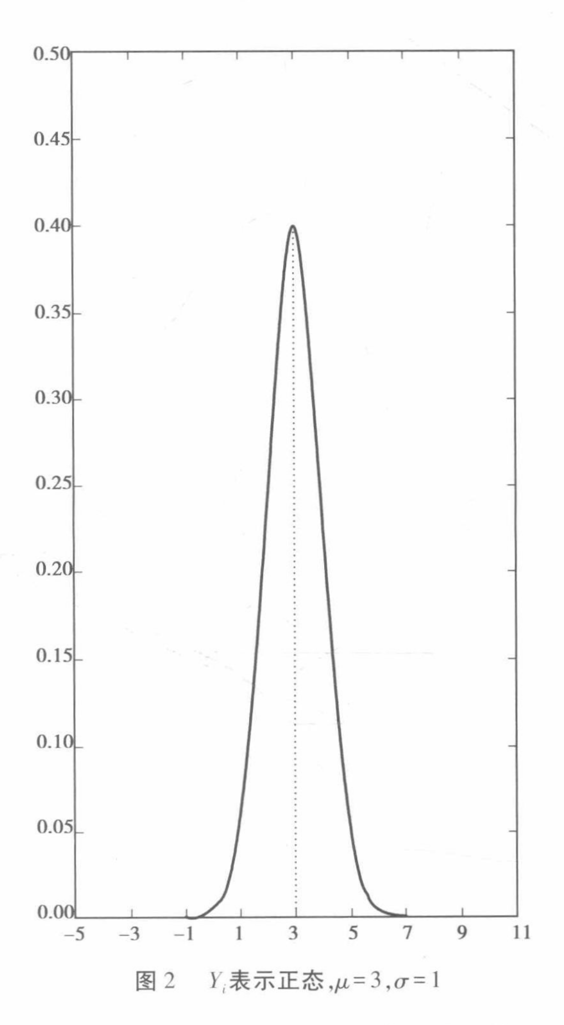


图 2 Y_i 表示正态，$\mu=3$，$\sigma=1$

在社会科学中，关于正态分布中心性的一个重要理论是中心极限定理【Central Limit Theorem】。通过证明在不考虑原始变量分布的情况下，对于大规模样本（或者更准确地说，n 趋向于无穷大）来说，变量的均值服从正态分布，那么中心极限定理就能将任意的概率函数（只要它的 μ 和 σ 是有限的）与正态分布关联起来。这个定理的一个特殊的应用是正态近似二项式【Binomial】分布。研究发现，在一系列的 n 次伯努利【Bernoulli】实验中，每次成功的概率为 p，近似于服从正态分布，其均值是 p，方差是 $\frac{p(1-p)}{n}$。

此外，根据最小二乘回归【Least Squares

Regression】中的误差【Error】项服从正态分布的假设【Assumption】,最小二乘估计值也服从正态分布,而这一特性对统计推论【Statistical Inference】很有用处。

标准正态分布

为了比较两个分别来自不同正态分布的变量的值,需要对其进行标准化。标准化将不同分布的均值和方差置于可比较的基点上,例如使得人们可以比较高级阿卡德语(Akkadian)和天体物理学导论的分数。

标准化程序是通过 $Z_i = \frac{Y_i - \mu}{\sigma}$ 来实现的,因此每个 Z_i 值表示 Y_i 与其均值的差,除以原始变量 Y_i 的标准差(σ)。因此,Z_i 的均值为 0,且标准差等于 1。

这样的话,标准正态分布的密度函数是

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

要对正态变量位于点 a 和 b[$P(a \leq Y_i \leq b)$]之间的概率进行估计,就需要估计正态密度函数下在点 a 和 b 之间的面积,或者换句话说,需要估计如下积分:

$$\int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \mathrm{d}y$$

不幸的是,由于并不存在积分的闭合表达式(closed-form expression),因此在此需要使用数值技术,在实践中则经常使用汇总了密度的统计表格。由于分布以 μ 为中心呈现对称性,因此需要仅在均值的一侧计算对称面积。或者为了方便起见,均值左侧的面积以均值右侧的面积计算方式来计算。图 3 用标准正态分布说明了这一点。

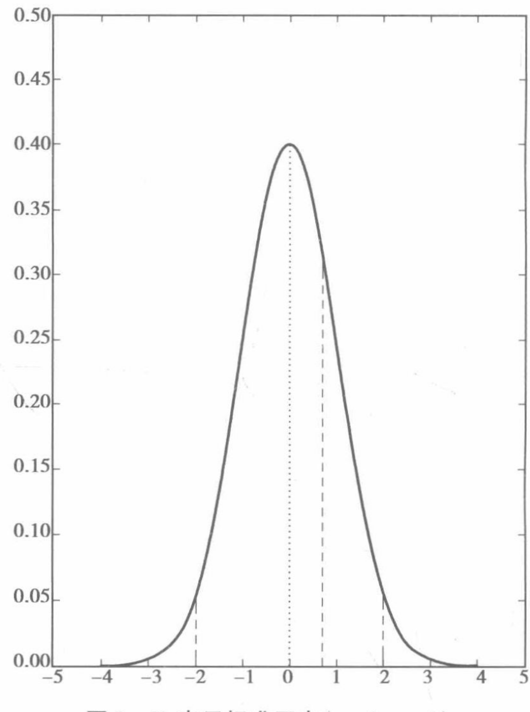


图 3　Y_i 表示标准正态($\mu=0, \sigma=1$)

$$\begin{aligned}
&P(a < Y_i < b)\\
&= P(Y_i < b) - P(Y_i < a)\\
&= P(Y_i < b) - P(Y_i > -a)\\
&= P(Y_i < b) - [1 - P(Y_i < -a)]\\
&= P(Y_i < b) + P(Y_i < -a) - 1
\end{aligned}$$

假设 $a=-1.5, b=0.8$,那么

$$\begin{aligned}
&P(-1.5 < Y_i < 0.8)\\
&= P(Y_i < 0.8) + P(Y_i < 1.5) - 1\\
&= 0.788\,1 + 0.933\,2 - 1\\
&= 0.721\,3
\end{aligned}$$

线性转换

社会科学家经常要对感兴趣的原始变量进行线性转换(Linear Transformation)来得到新变量。例如,研究者可以用美元或数千美元来测量收入,或者用公里和英里来测量距离。如果原始变量服从正态分

布，其线性转换也服从正态分布。进一步说，如果我们知道原始变量的均值和标准差，当两个变量是相关的，就很容易得到转换变量的参数值。如果 $Y_i \sim N$，那么 $E(Y)=\mu$ 且 $V(Y)=\sigma^2$，X 是 Y 的线性转换，即 $X_i=aY_i+b$，那么 $E(X_i)=E(aY_i+b)=aE(Y_i)+b=a\mu+b$，且 $V(X_i)=V(aY_i+b)=a^2V(Y_i)=a^2\sigma^2$。

多元正态分布

如果 Y_i 不是一个单独的变量，而是 N 个正态分布随机变量的向量，我们需要用多元分布来建构向量模型。含 N 个随机变量的多元正态分布可以表达为一个 $N\times1$ 的向量 μ 和 $N\times N$ 的方差协方差矩阵 Σ 组成的函数，即

$$f(y)=\frac{1}{\sqrt{|\Sigma|}(2\pi)^{-\frac{N}{2}}}e^{-\frac{1}{2}[(y_i-\mu)\Sigma^{-1}(y_i-\mu)]}$$

当 Y_i 只包括一个随机变量时，这个表达式可以缩减为单变量正态密度，其方差协方差矩阵是 σ^2。

如果变量彼此独立，那么方差协方差矩阵 Σ 就是对角矩阵（非对角线元素全部等于 0）。在这种特殊的情况下，单变量正态分布的乘积等同于多变量正态分布。

对数正态分布

对数正态分布有助于描述那些不能取负值的变量（例如收入或重量）。如果 Z_i 服从标准正态分布，可以得到 $Z_i=\frac{\ln Y_i-\mu}{\sigma}$，那么 Y_i 就服从均值为 μ、方差为 σ^2 的对数正态分布，表达式为 $Y_i=e^{\sigma Z_i+\mu}$。

发展历史

人们常常将正态分布与卡尔·高斯（1777—1855）联系起来。在 1809 年出版的一部著作中，高斯把正态密度与最小二乘法联系起来。但是，高斯引用了拉普拉斯 1774 年的著作，他从这里接触到了正态分布，事实上一些著作将这种分布称为高斯-拉普拉斯分布。历史学家还将正态分布的起源进一步向前追溯到棣莫弗 1733 年的一部著作（Stigler，1999，pp.284-285）。

实际上，正态分布可以用来描述统计过程的观念早在 18 世纪就出现了。在 19 世纪中的多数日子里，它们被称为“误差曲线”，因为最早用它们来描述测量误差的分布。直到 1870 年代，才正式引入了“正态分布”这个词。

——Orit Kedar

（李洋译　高勇校）

参考文献

Greene, W. H. (1993). *Econometric analysis*. Englewood Cliffs, NJ: Prentice Hall.

King, G. (1989). *Unifying political methodology*. New York: Cambridge University Press.

Stigler, S. M. (1986). *The history of statistics*. Cambridge, MA: Harvard University Press.

Stigler, S. M. (1999). *Statistics on the table*. Cambridge, MA: Harvard University Press.

Wackerly, D. D., Mendelhall, W., III, & Scheaffer, R. L. (1996). *Mathematical statistics with applications* (5th ed.). Belmont, CA: Duxbury.

正态化(Normalization)

在许多常见的统计模型中,我们假设变量 Y 的某个部分服从正态分布【Normal Distribution】。例如,在线形回归【Linear Regression】模型 $Y=\alpha+\beta X+\varepsilon$ 中,我们假设误差项 ε 服从正态分布。尽管稍微偏正态也是可以接受的,但是偏离正态的重尾分布可能会影响统计估计。在这种情况下,更好的处理办法是对 Y 进行转换,使得相关部分接近正态。这种变量转换被称为正态化。

如果变量 Y 的相关部分呈现重尾(偏态)分布,那么我们通常采用幂转换(power transformation)。顾名思义,它将 Y 转换为 Y 的 p 次幂(把 Y 转换成 Y^p)。大于 1 的幂可以降低负偏斜(negative skew),如二次转换 $Y^2(p=2)$。大于 0 小于 1 的幂可以降低正偏斜,如 Y 表示计数或频次时常用到的平方根转换 $\sqrt{Y}$ 或 $\sqrt{Y+1/2}\,(p=0.5)$。如果幂为 0,那么我们定义幂转换为 $\log(Y)$,它与较小的幂转换一样,也能降低正偏斜。负幂的作用等同于倒数 $1/Y$ 的正幂。如果可以合理地解释倒数的意义,就可以使用负幂指数,例如如果 Y 是发生率(单位时间的事件数),那么 $1/Y$ 就是两个事件间隔的时间。

总之,幂转换族可以表述为:

$$t(Y;p)=\begin{cases} Y^p & \text{当 } p\neq 0 \\ \ln(Y) & \text{当 } p=0 \end{cases}$$

幂转换假设 Y 是正的;如果 Y 是 0 或者负数,我们一般通过增加一个常数使 Y 成为正数。有一些正规方法可以估计要增加的常数和最接近正态性的幂指数 p(Box & Cox, 1964)。但是,最优幂和最优增加常数通常也只是一种大概的指引。

如果 Y 的相关部分的分布呈现两个重尾(超峰态【Kurtosis】),我们可以使用模转换(modulus transformation)(John & Draper, 1980),它分别对各尾进行修正的幂转换。小于 1 的非负幂指数 p 降低了峰态,大于 1 的幂指数提高了峰态。对最优幂指数 p 进行估计的正规方法也有许多种(John & Draper, 1980)。如果 Y 是以 0 为对称的,那么模转换会改变峰态而不引入偏态。如果 Y 不以 0 为中心,那么建议在进行模转换之前增加一个常数。

$$t(Y;p)=\begin{cases} \operatorname{sign}(Y)\left[\dfrac{(|Y|+1)^p-1}{p}\right] & \text{当 } p\neq 0, \\ \operatorname{sign}(Y)[\ln(Y)+1] & \text{当 } p=0 \end{cases}$$

如果 Y 是取值在 0~1 的比例,可以采用其他正态化方法:反正弦或角转换 $\mathrm{acrsin}(\sqrt{Y})$,Logit 或 Logistic 转换 $\ln\left(\dfrac{Y}{1-Y}\right)$,以及 Probit 转换 $\Phi^{-1}(Y)$,其中 Φ^{-1} 是累积标准正态密度函数的反函数。Logit 和 Probit 概率优于反正弦转换,后者已经逐渐没有人使用了。

即使是最好的转换也不一定能足够近似正态。此外,转换后的变量可能难以进行解释,从中推导出的结论难以应用到最初的、未经转换的变量上(Levin, Liukkonen, & Levine, 1996)。幸运的是,现代研究者除了正态化,还有其他处理方法。如果处理的是非正态数据,我们可以使用假设分布服从其他类型的广义线性模型【Generalized Linear Model】。我们还可以用"自由分布"统计量或非参数统计量来放宽假设。

除这里的定义之外,有时正态化这个术语也被用来指变量被标准化【Standardized】。有时也被用来指加入约束条件来确保联立方程【Simultaneous Equations】组可以识别

(如 Greene,1997)。

——Paul T. Von Hippel
(李洋译　高勇校)

参考文献

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society (B)*, *26*(2), 211-252.

Cook, R. D., & Weisberg, S. (1996). *Applied regression including computing and graphics*. New York: Wiley.

Greene, W. H. (1997). *Econometric analysis* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

John, N. R., & Draper, J.A. (1980). An alternative family of transformations.*Applied Statistics*, *29*(2), 190-197.

Levin, A., Liukkonen, J., & Levine, D. W. (1996). Equivalent inference using transformations. *Communications in Statistics, Theory and Methods*, *25*(5), 1059-1072.

Yeo, I.-K., & Johnson, R. (2000). A new family of power transformations to improve normality or symmetry.*Biometrika*, *87*,954-959.

NUD * IST

NUD * IST 是一种用来进行计算机辅助定性数据分析的软件程序。它能对定性数据进行编码和检索。它还支持扎根理论【Grounded Theory】应用,例如生成备忘录【Memos】。这个名称是非数字无结构数据索引、搜寻和理论化(Nonnumerical Unstructured Data Indexing, Searching, and Theorizing)的缩写。

——Alan Bryman
(李洋译　高勇校)

冗余参数(Nuisance Parameters)

冗余参数是一种在参数估计【Estimation】中研究者不感兴趣或者仅有次要兴趣的参数。例如,在一个包含正态分布随机变量的统计模型中,研究者的兴趣往往是估计变量的期望μ,而不是随机变量的方差σ^2,因此随机变量的方差就被视为冗余参数。

——Tim Futing Liao
(李洋译　高勇校)

虚无假设(Null Hypothesis)

在社会研究中,虚无假设是假设检验【Hypothesis Testing】的重要组成部分。在完成对社会活动的定量测量后,从理论推导出的假设就需要被检验,这时就会用到虚无假设。理论假设(或研究假设)认为数据会呈现某种模式,统计分析就是用来评估收集到的测量数据中的证据能够在多大程度上支持这种模式的存在。虚无假设就是认为发现的数据模式仅是由于偶然才出现的假设。

举例说明,从总体【Population】中随机

抽样【Random Sampling】出参与者，再随机分成两组。一组饮酒，另一组服用安慰剂（参加者并不知道自己属于哪个组）。然后对他们的驾驶表现（或其他相关技能）进行测量。研究假设是两组人员的表现水平不同。虚无假设是酒精没有影响，两组之间产生的任何差异都是由于随机的偶然性。如果虚无假设为真，我们期待两组的平均表现相同。但是，将参与者随机分配【Random Assignment】的两组总会在所测量的表现上有所差异，即便对两组人的安排完全相同。尽管如此，均值的差异越大，这种结果的可能性就越低。假设检验取决于我们得到测量的结果的随机概率（或几率）。

如果假定虚无假设为真（在本例中即酒精没有影响），可以计算出随机分配的两组中的测量结果呈现出某种数据模式的概率。通常，会用这个假设（虚无假设为真）来计算出（或估计出）某个统计量（如 t 比率【t Ratio】或 Chi-方【Chi-Square】）的抽样分布【Sampling Distribution】。然后这个分布会被用来和测量数据中计算得到的样本统计量进行比较。

理论而言，即便相关的处理没有效应（所以虚无假设为真），也完全可能仅仅由于偶然而在数据中呈现出某种数学模式，即使可能性极小。仅仅由于偶然而在结果中发现某一特定模式的概率越小，我们就越有信心认为有证据支持研究假设。一些统计方法会设定一个概率水平作为惯例，如果低于这个水平，程序就会作出拒绝虚无假设、接受研究假设的决策。另外，一些方法认为并不存在普遍适用的又好又快的法则，用某一组测量方式得到的研究证据，只有放在以类似方式探讨同样问题的其他研究的情境中，才能得到充分的评估（参见 MacDonald, 2002）。

——Martin Le Voi

（李洋译　高勇校）

*也可参见显著性检验【Significance Testing】、p 值【p Values】、假设【Hypothesis】。

参考文献

MacDonald, R. R. (2002). The incompleteness of probability models and the resultant implications for theories of statistical inference. *Understanding Statistics*, *1*(3), 167-189.

数值变量（Numeric Variable）

参见计量变量【Metric Variable】。

NVivo

NVivo 或 NUD＊IST Vivo 是一种用来进行计算机辅助定性数据分析的软件程序。依其设定，它能进行定性数据编码和检索。它还支持扎根理论【Grounded Theory】实践，例如生成备忘录【Memos】。它是 NUD＊IST 软件的一个变体（但并非其新版本），但经常被认为更适合细致分析，因此在定性研究【Qualitative Research】专业人员中非常流行。

——Alan Bryman

（李洋译　高勇校）

O

客观主义(Objectivism)

客观主义的观点认为,存在一个"永恒的、与历史无关的模型或框架,我们最终可以诉诸它,来确定理性、知识、真理、现实、善良或公正的本质"(Bernstein, 1983, p.8)。尽管长期受到社会科学中的现象学和诠释学传统的挑战,但是一直到1960年代,客观主义仍然是自然科学和社会科学中的默认规则。托马斯·库恩(Kuhn, 1970)1962年首次出版的《科学革命的结构》(*The Structure of Scientific Revolutions*),在自然科学领域(然后引申至社会科学)挑战了这一正统,宣称在科学中不存在线性进步;相反,多种被用来表达科学真理的思维和实践"范式"会在特定的时期共存。范式从属于偶然性的和戏剧性的革命,这些革命更容易用社会和心理因素进行解释,而不是关于世界的客观事实。事实上,竞争性范式是对立的,以致它们在认识论上是不可通约的(incommensurate)。

库恩的著作影响了社会科学中许多反客观主义者的观点(或者更准确地说是社会研究中,因为这些观点通常否认社会世界可以用科学方法来研究):相对主义【Relativism】、社会建构论【Constructionism】和主观主义。尽管一个反客观主义者有时会将所有这些对立观点混合在一起,但它们未必具有相同的含义。相对主义者会否认知识和/或道德可以根据一个没有嵌入在主观认识论或道德立场中的视角来判断,因此他们声称任何真理或道德都没有特权。主观主义者认为不存在位于主体的观念或建构之外的客观现实。虽然相似,但社会建构主义强调的是对我们所谓"现实"的社会建构。

客观主义,如同许多其他"主义"一样,往往在某种意义上说是一个"稻草人",现在很少有人会完全地信奉它或否定它。但是,有更多温和的继承者或者把客观性作为一种可实现的目标,或者把实在论作为一种本体论立场。前者的问题是:如果社会世界是由其参与者积极构建的,那么在追求社会世界的真理时,研究者还能在多大程度上保持客观性?实在论者则试图找到一种中间立场,他们接受社会世界是社会性地建构起来的观点,同时认为社会世界一旦形成就会作为一种独立于任何个体观念的客观现实而存在。

皮埃尔·布迪厄(Bourdieu,1977)在超越客观主义及其对立观点上作出了有影响力的尝试,他将客观主义与涂尔干、阿尔都塞和列维·斯特劳斯的结构主义【Structuralism】和社会理论中随之出现的主

体“消失”联系在一起。相应地，他认为常人方法论者【Ethnomethodology】和符号互动论者【Symbolic Interactionism】过度地关注行为主体的主观性而忽略了对它们起作用的客观结构。因此，他提出在建构社会现实的实践中，主观和客观因素之间存在着一种辩证关系。客观结构影响着日常互动，尽管主体会再生产和改变它们，但是它们所呈现的方式会影响它们如何被再生产和改变。

——Malcolm Williams

（叶鹏飞译　高勇校）

参考文献

Bernstein, R. (1983). *Beyond objectivism and relativism: Science, hermeneutics and praxis.* Oxford, UK: Basil Blackwell.

Bourdieu, P. (1977). *Outline of a theory of practice.* Cambridge, UK: Cambridge University Press.

Kuhn, T. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press. (Original work published 1962.)

客观性（Objectivity）

“客观性”一词的中心含义涉及的是，调查是否以一种使获得真实结论的可能性最大化的方式来进行的。当然，正如将会明确的，该词还有一些其他相关的含义。

这一概念的境遇在20世纪下半叶发生了巨大的改变。大多数研究者将其视为研究的基本要求，但也有一些人认为它仅仅是一种符合科学家利益的思想伪装，特别是那些来自西方国家的男性白人科学家。

“客观性”的不同含义可以通过区分它的形容词和副词形式来进行说明。混淆不清常常是因为将它们合并在一起。首先，“客观的（objective）”可以作为形容词来描述世界上的现象。从这个意义上说，它有时被用来表示已经存在的事物，而不是简单的表象或想象出来的虚构物。另外，“客观性（objectivity）”意味着一个事物属于“外部的”世界，而不属于主体的“内部的”、心理的世界。从这个意义上看，桌椅是客观的，而思维和感觉并非客观的，尽管它们可能都是真实的而不仅仅是表象。

“客观的”也可以作为形容词用来描述研究成果，这里它的意思等同于“真实的（true）”。有时这被称为“本体论的客观性”（Newell, 1986）。然而，既然是等同的，那么这种用法似乎是多余的，大多数情况下最好避免使用。

相比之下，第一段中认定的中心含义是副词性的，这里“客观性”一词是应用于调查过程和/或调查者。有时这被称为“程序上的客观性”，尽管这会导致得出错误结论，因为它意味着仅仅遵循一种方法。这种副词意义上的“客观性”是指，调查是否会因为外部因素的影响而产生偏倚【Biase】，尤其是研究者的个人动机和兴趣、社会背景、政治和宗教责任等。重要的是要注意，客观性并不要求这些因素对研究过程不产生任何影响，而仅仅要求它们不能使研究偏离了理性过程。

这表明副词意义上的“客观的”，是评估性的而不是事实性的：它涉及对调查过程或调查者的评判，依据是如何开展调查才能使产生真实调查结果的可能性最大化的理性标准。它并不关心调查过程中的全部偏差，而是关注与研究者相关的因素导致的系统误差。值得强调的是，在心理学意义上这些因素未必是主观的。

副词形式的客观性作为一种规约性理

想(regulative ideal),已经受到很多证据的挑战。一些评论者声称它永远无法达到,因为这意味着研究者没有任何背景特征。在此基础上,有人认为客观性的概念仅仅是一种修辞手法,研究者借此将他们自己的观点特权化。然而,正如前文所述,客观性并非要求没有背景特征,而仅仅是要求这些背景的所有系统负效应必须避免。此外,即使客观性真的永远无法完全达到——因为我们不能彻底避免这类负效应,或者至少知道我们已经避免了它们——但也并不是说追求达到客观性是不受欢迎的。有观点认为,调查越是接近这种理想,所得到的知识越有可能是可靠的。

有人提出特定类别的行为主体比其他人更能获得客观性。哈丁(Harding,1992)提出了一种观点。她认为,支持她所称呼的"强客观性(strong objectivity)",实际上是给予那些社会地位边缘化的人们的视角以特权,原因是这些人不会安于现状而不辨是非,从而更有能力获得关于社会的真知灼见。这一态度的基础是黑格尔学派和马克思主义的思想,意思是客观性是社会历史的产物,而不是个人取向的问题。

对作为理想的客观性的另一种批评,源于对它所依赖的理性和/或真理概念的否定,它们已经被视为虚构的或仅仅是修辞性的。这种批评属于怀疑主义和相对主义【Relativism】的一种形式,近来更受到后结构主义【Poststructuralism】的启发。

还有一种对作为规约性理想的客观性的独特批评认为,它鼓励了研究者在伦理上的不负责任。这是因为它要求在研究的实施过程中,不能将真理之外的重要价值作为目标的一部分。在批评意见中,真理和理性的观念没有被否定;要求很简单,调查必须包含对知识之外的其他目标的理性追求,这些其他目标——如社会正义——被假定为与真理有选择性的亲和关系,或者作为对真理的适当的重新定义。

客观性传统概念的捍卫者拒绝这些批评,他们认为"客观性"纯粹地意味着要避免来自不相关因素的偏见,没有人可以单独享有特权或者被阻止获取社会世界的知识;任何对知识可能性的否认都是自我反驳的(self-refuting);有充足的理由相信,不同的目标意味着理性行动的不同路线,因此无法同步进行。客观性的捍卫者声称,科学探索的成功是来自它在研究团体中的制度化,在研究中力求避免不相关事项造成的偏差,其重要性就如同在任何其他的理性行动中一样(参见 Rescher, 1997)。

——Martyn Hammersley

(叶鹏飞译　高勇校)

参考文献

Harding, S. (1992). After the neutrality ideal: Science, politics, and "strong objectivity." *Social Research*, *59*, 568-587.

Newell, R. W. (1986). *Objectivity, empiricism and truth*. London: Routledge and Kegan Paul.

Rescher, N. (1997). *Objectivity: The obligations of impersonal reason*. Notre Dame, IN: University of Notre Dame Press.

斜交旋转(Oblique Rotation)

参见旋转【Rotations】。

观察表(Observation Schedule)

在观察法研究中,研究者使用观察表,从而以一种有结构且系统的方式研究行为和互动(参见结构化观察【Structured Observation】)。观察表包括一些在行动和互动发生时对其进行分类和记录的预定系统,同时包括一系列用来指导观察和对事件进行归类的规则和标准。一个观察表通常要有一个时间维度,以便及时定位观察结果,同时要有与行为和互动的不同方面相关联的变量,每个变量由一组描述特定行为和互动的类别所构成。观察表是一种操作化的方式,研究者通过它将作为研究焦点的社会世界的某些方面进行操作化;它被用来收集可以用于分析的系统而可靠的数据。观察表中使用的类别和定义必须是明确的,并且能够被不同的观察者同等地使用。

观察表的类型

观察表有多种不同的类型,特别是围绕它们如何及时定位观察结果,它们是否提供有关所研究行为的持续期、频次、发生时间和结果等方面的数据。连续和准连续观察系统提供了一种对行为的持续、不间断的记录。这些系统是最完整的,但只能应用于有限的类别。例如,希尔萨姆和凯恩(Hilsum & Cane,1971)通过整天连续的观察来研究学校教师的工作,将其行为进行分类,比如"指导""评分"和"管理"。事件记录本质上是一个清单系统,标明特定类型的事件具体是何时发生的。这对于事件出现的频次是一种很好的测量。惠尔达和梅雷(Wheldall & Merrett,1989)采用这样的方法,根据对老师表扬或批评学生的行为进行简单计数,来比较课堂管理行为的不同方式。很多系统使用各种类型的时间抽样(time sampling)来构建观察结果,比如瞬时时间抽样以及部分间隔和完整间隔时间抽样。高尔顿、西蒙和克罗尔(Galton,Simon,Croll,1980)采用瞬时时间抽样研究了学校课堂;选出的学生与其老师之间的互动和行为,被精确地按照25秒的时间间隔进行编码。克罗尔(Croll,1986)、苏恩和阿里(Suen and Ary,1989)给出了关于时间抽样方法的进一步详细介绍。

设计观察表

在设计新的观察表之前,研究者应该考虑,在已经付诸使用的诸多系统中,有没有哪一种可以服务于他们的目标。即使现有的观察表都是不适合的,对于可比较的目标,也仍然值得考虑如何修改某些已经使用的部分,或者至少从另一个观察表中合并一些变量来作为基准测量。罗布森(Robson,1993,Chap.8)提出了发展观察系统的指南。其中包括这些标准:在研究情境下的易用性、对观察者推断力或判断力水平的要求足够低,以及对类别的明确界定。使用的类别应该是穷尽的,即所有遇到的行为都可以进行编码,而且应该是互斥的,即每个观察的行为都只能归入一个变量(当然也可能存在多个变量)的一种类别中。想要使用观察表的研究者可能会发现,最好是要以一种非结构化的方式展开观察,然后逐步把结构性引入观察过程中。这既有助于聚焦在他们所研究领域的概念化,也有助于确定更准确的操作化相关变量的方法。

——Paul Croll

(叶鹏飞译　高勇校)

参考文献

Croll, P. (1986). *Systematic classroom interaction.*

Lewes, UK: Falmer.

Galton, M., Simon, B., & Croll, P. (1980). *Inside the primary classroom.* London: Routledge and Kegan Paul.

Hilsum, S., & Cane, B. (1971). *The teacher's day.* Windsor, UK: National Foundation for Educational Research.

Robson, C. (1993). *Real world research.* Oxford, UK: Blackwell.

Suen, H. K., & Ary, D. (1989). *Analyzing quantitative behavioural observation data.* London: Lawrence Erlbaum.

Wheldall, K., & Merrett, F. (1989). *Positive teaching: The behavioural approach.* Birmingham, UK: Positive Products.

观察的类型(Observation, Types of)

观察是一种数据收集策略,包括系统地收集和调查在种种情境下发生的语言和非语言行为。这种数据收集的方法特别重要,尤其是在由于行为主体无法进行沟通(例如,同婴幼儿或神志不清的成年人)或无法提供足够详细的信息(例如,关于复杂的互动模式),而使通过自我报告难以获得相关信息的情况下。观察还被用于验证或扩展使用其他数据收集方法所获得的数据。非结构化观察和结构化观察都为研究者所使用。非结构化观察大多用于探索性和描述性研究。结构化观察【Structured Observation】用于感兴趣的行为是已知的情况,这种类型的观察一般要使用一个观察表【Observation Schedule】(如一张清单)。通常,要努力在尽可能自然的环境中对参与者进行观察。涉及观察的数据收集要获得参与者的知情同意【Informed Consent】。

观察有两种主要类型:亲身观察(参与观察【Participant Observation】或非参与观察【Nonparticipant Observation】)和录像记录。亲身观察是指研究者直接进行持续的观察,在聚焦情境的同时也关注个体行为,以理解某一行为或信念的意义。这一观察类型通常与民族志【Ethnography】有关。例如,参与观察对于研究敬老院环境中的隐私是一种很有用的方法。在此环境中,参与者与他人相处时所扮演的角色或参与的程度,可以是多样化的。在非参与观察中,研究者主要聚焦于观察这一任务,而最小化他们对环境中互动的参与度。在其他情况下,研究者可能更多的是一名参与者而不是观察者,因为在观察进行后,他们在研究的环境中也承担起一部分局内人的角色和职责。亲身观察的优点有:(a)经过一段时间,参与者适应了研究者的存在,当感兴趣的现象实际发生时,能够增加观察到的可能性;(b)研究者有机会与参与者进行互动,以澄清和扩大观察结果;(c)按照事件在日常生活中呈现出的本来状态来理解它们;(d)参与者言行之间的差异能够一目了然(Bogdewic, 1999)。在这种包括研究者心智在内的环境中实施并以现场"随笔"形式记录下来的观察,之后经过研究者的解释,扩展为详细的田野笔记【Fieldnotes】(Emerson, Fretz, & Shaw, 1995)。一般认为,参与者可以限制某些信息的公开并规定观察的界限。

在第二种观察类型中,摄像机被用来捕捉感兴趣的行为。录像记录的优势是,能为细致的微观分析或各种类型的分析提供一个永久的观察记录(Bottorff, 1994)。录像记录通常应用于定性的行为学和某些形式的谈话分析【Conversation Analysis】。例如,录像记录为研究医疗服务人员与病人之间的

互动模式提供了丰富的资料来源。当研究者需要研究行为的顺序、同时发生的行为和无法实时观察的非语言行为时，这种观察手段尤其有用。通过对录像记录的行为进行微观分析，可以获得对行为的深度描述。

——Joan L. Bottorff

（叶鹏飞译　高勇校）

参考文献

Angrosino, M. V., & Mays de Perez, K. (2000). Rethinking observation: From method to context. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 673-702). Thousand Oaks, CA: Sage.

Bogdewic, S. P. (1999). Participant observation. In B. J. Crabtree & W. L. Miller (Eds.), *Doing qualitative research* (pp. 47-69). Thousand Oaks, CA: Sage.

Bottorff, J. L. (1994). Using videotaped recordings in qualitative research. In J. M. Morse (Ed.), *Critical issues in qualitative research* (pp. 224-261). Newbury Park, CA: Sage.

Emerson, R. M., Fretz, R. I., & Shaw, L. L. (1995). *Writing ethnographic fieldnotes.* Chicago: University of Chicago Press.

观察研究（Observational Research）

观察是社会学家、文化人类学家和社会心理学家使用的基本研究技术之一。它可以独立作为一种资料收集方法或者与其他一种或多种资料收集手段搭配使用。观察研究的目的是在群体活动、交谈和互动等事件发生时进行记录，并确定这些事件对于参与者的意义。观察可以在研究者设计的实验室环境中进行，也可以在呈现选定行为自然轨迹的实地环境中进行。通过观察获得的资料，可以通过使用标准化的列表或研究指南进行编码量化，也可以通过开放式的叙事描述进行描绘。资料用文字或录音（音频和/或视频）形式记录，包括关于自然环境、行为、互动模式、意义（明确表达的或暗含的）和情感表达的数字化或叙事性描述。

研究者会依据采用的观察类型，力图达到以下一个或几个目标：(a)从被研究者的视角去看待事件；(b)关注看似平常的细节（即对任何事情都不会想当然）；(c)将观察的资料置于尽可能广泛的社会和历史背景中（避免对有限资料以偏概全）；(d)达到研究设计的最大灵活性（即使用尽可能多的方法收集资料，以完善基本的观察资料）；(e)对客观记录的资料进行仔细分析后，建构理论或解释性框架。

观察研究的类型

社会科学家进行观察研究有三种主要方式，这些方式之间有相当大的重叠。但为了进行定义，有必要对它们加以区分：(a)参与式观察【Participant Observation】，是一种深受田野工作者喜爱的研究方式，它要求与所研究的社区建立密切联系，并且研究者要长期融入这些社区的日常生活中；(b)反应性观察（reactive observation），与控制性环境有关，假设被研究者知晓研究的存在，但与研究者的互动仅仅是对研究设计中的元素作出响应；(c)隐蔽（非反应性）观察，是对没有觉察到自己被观察的人进行研究。

上述三种类型的观察研究，每一种都涉及三个特定性层次（levels of specificity）的步骤：

(a)描述性观察包括观察者对所有细节的注解和描述，观察者采取近乎孩童的立

场,消除一切偏见,不对任何事情想当然。这一过程产生大量的资料,其中一些会被证明是无关紧要的。这需要有一定的实地经验,才能够确定哪些资料是与参与者无关的。

(b)焦点式观察(focused observation)要求研究者仅仅考虑与当前手头上问题相关的材料,通常专注于定义明确的群体活动类型(如宗教仪式、政治选举)。

(c)选择性观察要求进一步探讨一般类型中的某个特殊形式(如入会仪式、市议会选举)。

观察研究中的客观性问题

社会科学家很早就意识到他们的现实存在会影响观察中的活动和环境,他们也努力坚持审慎的标准来收集和分析客观的数据,以便最小化观察者偏倚的影响。客观性是活动和环境中的参与者与观察者在发生事件上保持了一致性的结果;换句话说,研究者不能对所分析的事件强加理论上的或其他类型的先入之见。

20世纪上半叶,社会研究者采取了以下四种观察角色,其科学客观性程度被认为是依次增加:(a)完全的参与者或内部成员;(b)作为观察者的参与者(受过科学训练的内部成员);(c)作为参与者的观察者(成为共同体成员的局外人,这是现实中文化人类学家首选的一种立场);(d)完全的观察者(研究者与所观察的人或环境没有联系,隐蔽地进行观察,尽量减少与被观察者的互动)。

然而,完全客观性的理想让步于现代伦理中的知情同意【Informed Consent】。正如大多数伦理审查委员会所要求的,研究者要尽可能地告知有关研究目的和性质的问题,这几乎肯定要比过去的理想化情况涉及更多的互动。基于这样的假设,即如果一个人不能以某种方式成为一名成员,就不可能起到伦理研究者的作用,现代研究者更倾向于根据在观察环境中的隶属程度来塑造角色。成员的角色可能包括:(a)外部的成员研究者(那些使用内部成员的视角,而没有参与到团队成员核心行动中的研究者);(b)活跃的成员研究者(那些参与到团队核心活动中的研究者,有时甚至承担领导责任,但并不一定完全认可团队的价值观和目标);(c)完全的成员研究者(那些研究他们已经身处其中的环境,或者在他们进行研究时已经完全被吸纳为成员的研究者)。但是,最后一类研究者仍然力图确保他们的内部身份不会以一种不自然的方式改变互动的流程——他们作为内部成员的身份可能会使他们能够分辨事件的自然流程和不自然流程。

当代的观察研究者更愿意在他们所研究的团体中拥有一个已经存在的成员身份(或者建立起一个新的成员身份)。他们认识到,在一个既定的社会情境中探寻"真理",要想完全过滤掉研究者自身的观念,既不可行也不合适。换句话说,以系统的严密性来使结果尽可能客观呈现出来的传统目标,已经被一种新的态度所取代,即"真理"可能存在多种形式,除此之外,观察结果是指特定研究者和特定对象在特定时点上交会的记录,而不是对某种永恒的、抽象的社会现实的反映(这种态度转变在观察性的田野研究者中更加明显;而实验研究者仍然选择控制额外变量,继续追求严格的客观性)。

不同于实验室研究人员,田野研究者也重新定义他们研究的人群。他们倾向于避免对这些研究对象使用传统术语,如"被试者",而更喜欢在研究中代之以"伙伴""合作者"或参与者。"被试者"术语看上去有点家长式作风,显现出一种令人不悦的图像——无所不知的科学家以一种超然的、控制的、抽象的(以及含蓄的、不切实际的)方式将人类置于显微镜下。因为

伦理规范要求告知"被试者"他们的权利和研究项目的性质与范围，所以这些人不再是被一个疏远的权威人物所观察的被动行动者，他们的价值观、信念和想法必定充溢于研究中。

观察研究中的效度和信度问题

社会研究者对在特定环境中对特定人群进行观察，就能够对他们的文化进行普遍性的描述，产生了越来越多的质疑。观察研究者通常会使用几种技术来提高资料的效度。观察研究者对研究总体之外的推论非常谨慎，但仍然对大范围的模式抱有兴趣，他们可以依据一个研究者团队在不同情境中运用标准化观察表（参见结构化观察【Structured Observation】）收集资料所形成的组合图像，而不是从单一的观察场所出发进行推论。有时，这意味着团队成员应该在年龄、性别和其他个人背景特征上具有多样性，从而确保这些因素不会在不经意间影响研究者的认知。这一方法的经典案例，是约翰·怀廷及其同事在1950年代进行的"六种文化研究"，提供了在子女抚养实践上跨文化的有效资料。新近的一个例子是由史蒂芬·凯斯特、米歇尔·克拉兹和劳里·普赖斯为国家药物滥用研究所所做的、在美国20个地方进行的药物注射者针头卫生研究。

资料的信度可以通过三角检验【Triangulation】加以提高，三角检验包含使用一系列资料收集方法进行的观察，以期获得合理的综合性结论。通过不同的资料收集技术得到相同结论的事实，使资料更有可能是可靠的。

几种社会研究理论与实践学派中的观察研究

观察研究深受社会科学中几种主要理论或概念传统追随者的喜爱。例如，形式社会学学派（乔里·西梅尔是早期代表）聚焦于社会互动和社会关系的形式或结构，而不是这些关系的特定内容。更多现代的形式社会学家，如曼弗雷德·库恩和卡尔·库奇，将社会现实看成自我与他人或多或少通过相互联系的惯例形式进行互动而建构出来的。该学派的支持者主张使用视频技术来记录观察资料；那些具有文化人类学倾向的研究者喜欢在现场录制，而其他人则喜欢将观察对象带到专门设置的可录制环境中。此类研究倾向于聚焦人们建立和维持关系的方式，尽管改进后的技术方法也已被应用于研究大型政府机构的运作。

与欧文·戈夫曼的研究相关的拟剧社会学，集中关注人们如何在社会环境中呈现自己，以及事实上如何在"观众"即社会群体面前表演他们的自我。戈夫曼的方法，尽管有时被批评为缺乏系统性并因此过于主观，但它能够使研究者捕获人们在不同社会层面上的一系列行为。戈夫曼的追随者如斯宾塞·卡希尔，尽管他在公共和半公共的场合使用了戈夫曼的无结构式记录方法（现场做笔记，之后再进行分类，而不是把观察到的行为立即分类到预先编好的类别中），他还是尝试着在方法中引入一定程度的系统性，尤其是利用研究团队。与之相比，劳德·汉弗莱斯试图构建一个系统性的方法来捕捉拟剧的细节。在他关于公共卫生间中同性恋相遇的研究中，他按照观察表记录了观察结果，然而这种方法由于可以识别参与者的身份（他们实施的是私密行为，可能本来不想让别人如此近距离地观察自己），引起了对此类研究的伦理问题的批评关注。

自我观察（或自传式民族志【Autoethnography】）是这样一种技术，将观察的单元从公共或半公共场合下的群体转移到个体及她或他的亲密关系中。通常，谈论的个体就是研究者自身。这一方法有时被称为存在主义社会学，其哲学根源可以追溯到社会学家威廉·狄尔泰，他提倡一种理

解(verstehen)(通过移情来理解)的立场。卡罗琳·埃利斯是近代一名重要的实践者,她因其有影响力的关于伴侣去世的个人陈述及其处理悲伤的方法而为人所知。虽然这个案例是社会学家本身,但是研究的主题具有普遍性,她所展示出来的应变代表了她所身处其中的文化和社会。该方法在概念上与反思人类学学派相关,参与式观察者放弃严格的客观性态度,而将自身放置于他们所描述的文化叙事中。文森特·克拉潘扎诺和彼得·威尔逊是这一方法的两位杰出实践者,他们通过与被调查者的个人关系这一视角来创建民族志(分别在摩洛哥的一个社区和加勒比海的一个岛屿)。

也许没有哪个社会研究学派能像常人方法论【Ethnomethodology】那样重视观察技术。该方法的支持者认为,社会生活是依据潜意识过程来构建的,而这个潜意识过程涉及的是人们习以为常的微小姿态,因此无法通过向人们提问来获取。常人方法论学者通常喜欢运用复杂的符号系统来分析微小细密的资料,这是他们与认知人类学家所共享的一种方式,后者已经将注意力转向注重谈话的重复、停顿和语调的话语分析。社会学家道格拉斯·梅纳德使用这一技术来研究庭审中的语言,人类学家迈克尔·阿加用它来研究康复计划中的吸毒者行为。非语言交流也可以通过这种方法进行记录,比如由爱德华·霍尔率先提出的空间关系学方法(the methodology of proxemics)(关于特定文化中人们利用空间的方式以及人们在空间中如何安排自己——例如房屋中家具的摆放位置,候机室中人们的坐姿——的研究,表达了人际关系的文化意义),以及由伯德惠斯特利提出的身势学方法(the methodology of kinesics)(这一学术解释在通俗心理学中被称为"身体语言")。

观察研究的伦理维度

尽管如前文所述,观察研究者确实意识到与研究总体中的成员加强互动的必要性,但是他们并不认为他们从事了有害的侵入性工作。被界定为"公共的"观察研究(比如研究电梯中人们确定自身位置的方式)可以免除机构审查,然而互动极少的纯粹的观察研究案例变得越来越少。互动越多,产生负面影响的可能性就越大,很多情况无法在项目开始时充分预料到。因此,对于观察研究者来说,最重要的是,无论互动的程度如何,都要尽可能地对被研究者开放,尽可能地公开项目的方法和目标。

——Michael V. Angrosino

(叶鹏飞译 高勇校)

参考文献

Adler, P. A., & Adler, P. (1994). Observational techniques. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 377-392). Thousand Oaks, CA: Sage.

Agar, M. (1973). *Ripping and running.* New York: Academic Press.

Angrosino, M. V., & Pérez, K. (2000). Rethinking observation: From method to context. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 73-702). Thousand Oaks, CA: Sage.

Cahill, S. (1987). Children and civility: Ceremonial deviance and the acquisition of ritual competence. *Social Psychology Quarterly*, *50*, 312-321.

Crapanzano, V. (1980). *Tuhami: Portrait of a Moroccan.* Chicago: University of Chicago Press.

Ellis, C. (1995). *Final negotiations: A story of love, loss, and chronic illness.* Philadelphia: Temple University Press.

Gold, R. L. (1958). Roles in sociological field observation. *Social Forces*, *36*, 217-223.

Humphreys, L. (1975). *Tearoom trade: Impersonal*

sex in public places. Chicago: Aldine.

Maynard, D. W. (1984). *Inside plea bargaining: The language of negotiation.* New York: Plenum.

Price, L. J. (2002). Carrying out a structured observation. In M. V. Angrosino (Ed.), *Doing cultural anthropology* (pp. 107-114). Prospect Heights, IL: Waveland.

Schensul, J. J., & LeCompte, M. D. (Eds.). (1999). *Ethnographer's toolkit.* Walnut Creek, CA: AltaMira.

Whiting, J. W., Child, I. L., & Lambert, W. W. (1966). *Field guide for the study of socialization.* New York: Wiley.

Wilson, P. J. (1974). *Oscar: An inquiry into the nature of sanity.* New York: Vintage.

观察频次(Observed Frequencies)

频次可以被简单地定义为我们所测量的事件发生的次数。这个事件就是我们有兴趣获取信息的变量【Variable】。请看表 1 中的数据。

表 1　一周时间内的停车数

一周中的天数	停车数
星期一	10
星期二	9
星期三	10
星期四	8
星期五	1
星期六	3
星期日	2

我们测量的事件/变量是停车场的使用情况。于是,星期一的频次(即停车数)是 10,星期二的频次是 9,星期三的频次是 10,以此类推。

观察频次被定义为我们所观察到的测量事件发生的实际次数。这是否意味着我们必须待在停车场并记录每一辆停放的汽车?在有些案例中的确如此,我们可以亲临现场测量事件并统计它的频次。但在现实中,我们倾向于使用研究工具(例如问卷【Questionnaire】)来获取所测量事件/变量的详细资料。例如,可以请求停车场管理员填答一份关于停车场一周中每天使用情况的问卷,利用这些信息编制出表 1。

我们可以通过考虑相对频次【Relative Frequency】更详细地说明问题,相对频次是由观察频次转化而来的。相对频次是我们所要调查的现象发生的案例数占研究中全部案例数的比例。

例如,一周时间内,停车场中的停车总数是 43。可以通过下面的公式来计算本周中每天的相对频次:

$$相对频次 = \frac{某个特定日子里的停车数}{一周时间内停车总数}$$

这样,星期一有 10 辆车停放,一周时间内的停车总数是 43,那么星期一的相对频次就是:

$$相对频次(星期一) = \frac{10}{43} = 0.23$$

根据这个结果,我们就能把星期一停车场使用情况的相对频次表述为 0.23。我们可以重复这样的算法,将一周中其他天数里的相对频次也计算出来。

观察频次不能与期望频次相混淆。尽管二者由于观察同一事件/变量而彼此关联,但是在它们的计算和解释上存在很大

不同。

期望频次不是观察频次，尽管它可能是以后者为基础。期望频次是我们期望发生的，而不是实际发生的次数。期望频次通常建立在一些以前的信息基础上，即一个研究者已经获得的、计算出的或者记录下来的信息，通过这些，研究者试图预测将要发生的事件。例如，我们可以观察四个连续的星期一分别有多少辆汽车停放，从这个数据出发，来预测第五个星期一期望看到的停车数。

——Paul E. Pye

（叶鹏飞译　高勇校）

参考文献

Kerr, A. W., Hall, H. K., & Kozub, A. (2002). Descriptive statistics. Chapter 2 in *Doing statistics with SPSS*. London: Sage.

Hinton, P. R. (1996). *Statistics explained: A guide for social science students*. London: Routledge.

观察者偏倚(Observer Bias)

尽管运用观察法的社会研究者在记录和分析资料时会力求客观性【Objectivity】，他们还是意识到，源于观察自身性质的偏倚可能会损害他们的研究工作。因此，研究者需要有意识地察觉这些可能性，并采取措施避免它们。

观察者偏倚可能来源于研究者怀有的无意识的假设或偏见。在某些情况下，这些偏见表现为种族中心主义——未加反思地认为自身文化中的价值观、态度和习惯才是规范的，导致无法看到更无法理解那些与这种规范不相符的行为。女性主义学者提醒研究者注意那些建立在性别和性别角色中暗含的权力关系基础上的不言而喻的假设。其他批评者指出，种族中心主义根源于社会阶级、宗教背景乃至年代（即"代沟"），所有这些都可能导致研究者对于那些不同于自身的人们的生活会错失很多重要的方面。

种族中心主义的观察者偏倚的影响，通常可以通过严格而持续的努力来加以减小，这种努力致力于留意可能导致错误理解的无意识假设。有点更加微妙——也因此更难克服的——是由于人们知道自己在被观察时会倾向于改变自身行为而造成的偏倚，即使进行观察的研究者尽全力来将种族中心主义的影响最小化。

传统的文化人类学是以在所研究的社区中建立长期关系为基础的，这样社区成员就会忘记他们一直在被人观察，而观察者也不再是一个陌生人。但是，即使是人类学家也不再可能总能安排长期的田野研究。因此，有必要设计其他方法来减轻短期研究中的观察者影响；例如，研究者可以通过由某个德高望重的人士介绍到社区，来增加在研究群体中让人感到舒适和信任的程度。也有必要进行三角检验【Triangulation】——从其他来源的观察资料中获得多种印象的汇总，更好地保证来自任何一个研究者的无意识暗示都不至于歪曲结果。

大多数当代社会研究者认识到，他们无法预知自己对研究的社区可能产生的全部潜在影响，并且他们认为追求严格的中立态度是不切实际的。研究者某些方面的自我呈现可以很容易地加以改变，以便他或她在研究社区中显得不那么扎眼（例如着装的样式、语调）。其他因素（例如种族、性别、年龄）则无法改变，因此在伦理上，当研究者报告结果时，他们需要坦白地公开这些因素，以便读者能够自己判断研究者是否严重干

扰了研究人群对研究情境作出的响应。

——Michael V. Angrosino

(叶鹏飞译 高勇校)

参考文献

Gouldner, A. W. (1962). Anti-minotaur: The myth of a valuefree sociology. *Social Problems*, 9, 199-213.

Schensul, S. L., Schensul, J. J., & LeCompte, M. D. (1999). *Essential ethnographic methods: Observations, interviews, and questionnaires.* Walnut Creek, CA: AltaMira.

发生比(Odds)

一种状态出现(如某人是一名计算机使用者)或一个事件发生(如某人实施了一项犯罪)的发生比,是这一状态或事件发生的概率与其不发生概率的比值。如果我们用符号 $P(Y)$ 表示某一状态或事件的概率,那么 Y 的发生比是 $\text{odds}(Y) = P(Y)/[1-P(Y)]$。如果事件或状态是不可能发生的,那么概率为 0,发生比同样为 0。如果事件或状态的概率小于 0.5,那么发生比的范围为 0~1;如果概率恰好是 0.5,那么发生比为 1;如果概率大于 0.5,那么发生比也将大于 1;如果状态或事件是确定发生的,那么 $P(Y)=1$, $\text{odds}(Y)=1/0$,是一个正的无穷大数值。关于某一结果发生比的考虑,或者两项可选结果发生比的比较,在赌博中是常见的,这也在一定程度上推动了概率论和统计的早期发展(Stigler, 1986, pp.62-63)。

发生比是两个"对立"概率的比值,第一个概率等于 1 减去第二个概率。涉及两个概率的另一个相关的比值是相对风险或风险比,它是两个不同的但未必对立的概率的比值。例如,如果男性使用计算机的概率是 0.60,那么男性使用计算机的发生比 $\text{odds}(Y)$ 为 $(0.60)/(1-0.60)$,等于 1.5;如果女性使用计算机的概率是 0.45,那么女性使用计算机的发生比是 $(0.45)/(1-0.45)$,等于 0.82。然后,男性使用计算机的概率(或风险)与女性的比值就是相对风险,为 0.60/0.45,等于 1.33。换句话说,男性使用计算机的可能性是女性的 1.33 倍。第二个相关的比值是优势比【Odds Ratio】,是指两个不同的但并不对立的发生比的比值。在上述例子中,男性成为计算机使用者与女性相对照的优势比是:优势比=(男性使用计算机的发生比)/(女性使用计算机的发生比)= 1.5/0.82 = 1.83。换句话说,男性使用计算机的发生比是女性的 1.83 倍或高出 83%。要注意相对风险和发生比并不一样。一个常见的错误是计算优势比后,把它解释成了相对风险,例如在上面的例子中,男性使用计算机的可能性是女性的 1.83 倍(而如上面所示,男性成为计算机使用者的可能性仅仅是女性的 1.33 倍)。明确(a)概率或概率比与(b)发生比或优势比之间的区别非常重要。

——Scott Menard

(叶鹏飞译 高勇校)

参考文献

Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900.* Cambridge, MA: Belknap Press and Harvard University Press.

优势比(Odds Ratio)

优势比是两个发生比【Odds】相除。发生比是由两个代表不同事件的数字构成的(通常是发生事件相对未发生事件),例如胜利相对失败的发生比,投票民主党候选人相对投票共和党候选人(或非民主党候选人)的发生比。就其本身而言,优势比和比率是不同的,比率是把某个总体中事件的发生次数标准化到一个基础上,比如百分之几或千分之几。在分类数据分析【Categorical Data Analysis】中,研究数据中的变量关系或一致性,优势比是非常实用的统计工具。

乌迪·耶尔是第一位应用优势比来测量列联表【Contingency Table】中变量关系的著名统计学家。优势比和对数优势比是对数线性模型【Log-Linear Model】中的常用方法。为展示优势比的应用,让我们来考察表1中由格林伍德和乌迪·耶尔提供的典范性数据,数据来自他们1915年关于伤寒疫苗接种影响的经典研究。

表1 伤寒疫苗接种的数据

	伤寒:没有	伤寒:有	总 计
疫苗接种:是	6 759	56	6 815
疫苗接种:否	11 396	272	11 668
总 计	18 155	328	18 483

数据来源:Greenwood and Yule (1915)。

这是一个典型的2×2表格,列表示两种结果,行表示暴露变量的两种类别。一般情况下,列可以用来表示任何响应变量或因变量,行可以表示任何解释变量或自变量;当然,行和列只能包括两个我们希望研究其相互关系的变量。单元格可以用 A,B,C,D 来表示(参见表2)。

表2 一个2×2表格的结构

	变量1 类别1	变量1 类别2	总 计
变量2 类别1	A	B	$A+B$
变量2 类别2	C	D	$C+D$
总计	$A+C$	$B+D$	$A+B+C+D$

在一个2×2表格中,有三种不同的方式构造优势比(OR):

$$OR_1=\frac{A/C}{D/B},OR_2=\frac{A/B}{D/C},\text{以及 } OR_3=\frac{A/B}{C/D}$$

要研究两个定类变量之间的关系,通常采用OR_3。这一优势比也被称为交叉相乘比,因为 $OR=\frac{A/B}{C/D}=\frac{AD}{BC}$。从这里开始,我们将省略它的下标,因为这是我们唯一要详细考察的优势比。回到伤寒疫苗接种的数据,可以得出优势比或交叉相乘比为:

$$OR=\frac{(6\,759)(272)}{(56)(11\,396)}=2.881$$

该比值表示在疫苗接种和伤寒病发生率之间存在关系。对于那些接种疫苗的人,预防伤寒病的发生比是那些没有接种疫苗者的近3倍。像这样的统计资料,可以为政府机构的决策,如疫苗接种计划提供一些证据。

注意,耶尔的 Q 是优势比的一个简单函数,取值范围是-1~1:

$$Q=\frac{\text{OR}-1}{\text{OR}+1}=\frac{2.881-1}{2.881+1}=0.485$$

因为 Q 的数值接近 0.5(0 表示相互无关),说明疫苗接种和预防伤寒病之间存在中等正相关。

通常,使用(自然)对数优势比非常便捷,它具有两个有用的性质。对数优势比用 0 这一取值表示两个变量无关,而优势比是用 1 来表示。用 0 表示无关可能更直观,对一些研究者更有吸引力。而且,人们可以计算对数优势比的渐进标准误,

$$\sigma_{\log(\text{OR})}=\sqrt{\frac{1}{A}+\frac{1}{B}+\frac{1}{C}+\frac{1}{D}}$$

$$=\sqrt{\frac{1}{6\ 759}+\frac{1}{56}+\frac{1}{11\ 396}+\frac{1}{272}}=0.148$$

并按照标准的正态分布计算出相关的 Z 值:

$$Z=\frac{\log(\text{OR})}{\sigma_{\log(\text{OR})}}=\frac{0.460}{0.148}=3.108$$

这个 Z 的数值说明结果极为显著(即发生的发生比基本上不可能是因为偶然性而被观测到的)。我们可以更进一步,运用得出的 Z 值为对数优势比构造置信区间(或通过反对数函数构造优势比的置信区间)。

——Tim Futing Liao

(叶鹏飞译　高勇校)

参考文献

Greenwood, M. & Yule, G. U. (1915). The statistics of antityphoid and anticholera inoculations, and the interpretation of such statistics in general. *Proceedings of the Royal Society of Medicine 8*(part 2), 113-194.

Rudas, T. (1997). *Odds ratios in the analysis of contingency tables*. Thousand Oaks, CA: Sage.

官方统计(Official Statistics)

代表统治者的行政官员为了征税或征兵的目的进行统计,是很多古代文明的基本特征。但是以更广泛的方式系统化地汇编各种事实和数字以供国家使用,仅仅是在启蒙时代、科学革命和欧洲民族主义增长之后。如今,现代国家收集和公布关于人口、贸易、财政、住房、健康、消费、休闲活动以及很多其他社会现象的一整套数据【Data】,既反映当前的成绩,又可以为未来的政策制定提供指导。此外,诸如欧盟委员会、联合国及其附属机构、世界卫生组织和世界银行之类的超国家组织,也都在监控它们的管辖领域,以便形成关于区域和全球发展状况的广泛图景。

官方统计的主要类型

有些官方统计数据是在行政管理过程中例行产生的,并且为国家提供这些信息通常是由法律进行规定的。税收详情、出生人数、结婚数、离婚数、死亡数和某种法定传染病数,都是通过这种方式进行记录,类似的还有关于就业和失业、犯罪以及教育方面的数据。在很多国家,对这些收集数据的统计是定期公布的。

尽管这些统计资料似乎包含了全部记录,但是它们仍然常受质疑,原因是它们仅仅提供了它们所宣称覆盖领域的不完整图

像。最容易受到这类批评攻击的是关于犯罪、就业和失业的官方数字。另一种收集犯罪信息的方式,比如受害人调查【Survey】,显示出在人们遭受的犯罪程度与发布的犯罪记录之间存在很大差距。这个相差部分被称为"黑数(dark figure)",部分来源于很多犯罪没有被报告的事实,原因多种多样,包括受害人或证人对能否采取有效行动抱有怀疑态度,害怕犯罪者的打击报复,或者仅仅是因为漠不关心(Coleman & Moynihan, 1996)。

同样,官方就业统计没有考虑那些非正式或非法经济中工作的劳动者,失业统计遗漏了那些在工作稀缺的萧条市场中,对找工作失去信心和退出劳动力市场的人。由于失业率具有政治敏锐性,所以政府会采取一些行政手段进一步减少报告的失业数,比如将长期失业者和其他没有资格获得失业补偿的人排除在外。

最广泛并且最早建立的官方统计形式是人口普查【Census】的结果,这是有关一国之中所有国民信息的概要。美国人口普查局每隔10年进行一次全国人口普查。尽管人口普查被认为是关于一个国家全部人口的最全面的信息来源,但是它所注重的广度也使它在更详细的调查方面不够充分。

近几十年来,政府调查有所扩展,经常会针对某些特定的主题范围,调查人口普查难免会留下的空白部分。由于成本原因,用这种方式产生的很多类型的官方统计,并没有立足于所有可能收集到的数据,而是利用由一个大规模的代表性样本【Representative Sample】所组成的子集,然后将得出的结论加以推广,用来估计整个总体【Population】。使用访谈【Interviewing】或邮寄问卷【Questionnaire】收集数据,通常可以获得较高的回答率,能够增加对推论的信心。

英国这些调查的经典案例是最早于1950年代开展的家庭消费调查以及1970年代的综合住户调查。正如其名称所示,后者是一个被众多研究者所使用的通用性的数据收集工具,并且具有足够的灵活性,可以根据社会趋势和政策重点的变化,通过引入新的问题进行修订。其他主要的调查还包括同期群研究——例如跟踪1%英国人口样本的纵贯性调查(longitudinal survey)、劳动力调查,以及英国家户定群调查(British household panel survey)。

后两个研究是国家资助调查的案例,在设计上与其他国家的调查相匹配,因此其结果可用于国际性统计报告中。这些跨国统计越来越规范化,并且标准化测量方法的采用,不仅提供了可比性,也能防止个别政府的故意篡改。然而问题依然存在,当统计涉及政治敏感议题时,比如政治难民,就会影响其传播(Singleton, 1999)。

官方统计使用渠道的变化

在20世纪最后几十年期间,长期以来收集和编纂官方统计数据的行为完全改变了意义。虽然人口普查长期使用打卡处理(punched card processing)(美国1880年开始,英国1911年开始)来自动记录,但在20世纪后期,由于个人计算机处理能力的巨大突破和快速普及,普通人可以获得的计算能力大大扩展,同时万维网获得爆发式增长,可以存储海量统计信息的数据库急剧增加。这些并行发展的过程,意味着没有必要再完全依靠政府或其他官方机构公布报表,个人现在都可以访问数据集并作出他们自己的解释。

这些技术发展并不是发生在政治真空中。政府,特别是经济合作与发展组织(OECD)国家,采取民主化议程来使其公民可以获得更多的信息,以便能够更全面地了解社会变化,更积极地参与到国家治理的讨论中。目前,很多政府和组织对他们网站上

的数据库提供了自由入口。同时,国际性统计资料也有欧盟统计局这样的超国家组织进行整理和提供使用。然而应该认识到,数据产生和发布的方式也对这种可以自由获得看似可信和客观信息的数据天堂施加了隐性的限制。

官方统计的政治环境

1980年代,英国政府试图控制某些研究成果的发布,比如揭示富人与穷人死亡率差异不断增大的成果,还有类似前文提到的失业率统计等问题,这种企图提醒我们,某些统计数据具有高度的政治敏感性(Townsend, Davidson, & Whitehead, 1988)。我们不应该忘记,尽管某些公布的报表被称为"官方的",但这并不能保证它们的准确性或公正性,也不能保证没有出现过隐蔽的审查或施压。在世界范围内,公正性都是所有官方统计资料的使用者所关心的问题;在英国,尽管政府承诺信息自由和独立的国家统计服务,这种争论也在继续。

——Will Guy

(叶鹏飞译　高勇校)

参考文献

Coleman, C., & Moynihan, J. (1996). *Understanding crime data: Haunted by the dark figure*. Buckingham, UK: Open University Press.

Singleton, A. (1999). Measuring international migration: A case study of European cross-national comparisons. In D. Dorling & S. Simpson (Eds.), *Statistics in society: The arithmetic of politics* (pp. 148-158). London: Arnold.

Townsend, P., Davidson, N., & Whitehead, M. (1988). *Inequalities in health: The Black report and the health divide*. Harmondsworth: Penguin.

Ogive

参见累积频次图【Cumulative Frequency Polygon】。

OLS

参见最小二乘【Ordinary Least Squares】。

Omega 方(Omega Squared, ω^2)

Omega 方(ω^2)是对效应强度(或大小)的测量。效应强度有多种测量方法,它们是与不同的数据测量方法相对应的(例如,相关系数【Correlation Coefficient】就是一种效应强度的测量)。ω^2 与运用单因素或析因方差分析【Analysis of Variance】的实验研究相关。它是方差分析中与 R 方【R-Squared】(R^2)(R^2 是回归和相关中效应强度的测量)有关的一系列效应强度测量方法中的一种。效应的强度具有重要的科学意义,因为它给出了一个比较性的测量,说明一项处理(在实验研究中)会在多大程度上影响行为

测量。一个效应值为零,意味着这项处理没有效果;与 R^2 有关的效应强度测量的最大值均为1(比如相关系数)。

表1是一个单因素(无关测量)方差分析的结果。

表1 一个单因素方差分析

来源	自由度	平方和	均方	F 值
实验处理	k	a	x	$\frac{x}{y}$
误差	$n-k-1$	b	y	
总和	$n-1$	c		

注:k 是处理组的对象数,n 是全部测量对象数(在社会科学研究中,通常是指参与者)。

表2 一个单因素方差分析的结果示例

来源	自由度	平方和	均方	F 值
实验处理	3	600	200	2
误差	20	2 000	100	
总和	23	2 600		

对于固定处理效应,ω^2 的数学定义是

$$\omega^2 = \frac{a - (k-1)y}{c + y}$$

假设一个单因素方差分析的结果如表2所示,那么

$$\omega^2 = \frac{600 - (3-1) \times 100}{2\,600 + 100}$$

随机效应的定义有所不同,因为它的方差成分计算与固定效应大不相同。

更一般的公式是

$$\omega^2_{效率} = \frac{\hat{\sigma}^2_{效率}}{\hat{\sigma}^2_{总和}}$$

其中,$\hat{\sigma}^2_{效率}$ 是一个效应的方差成分【Variance Component】。假如方差成分的计算是恰当的,那么这个公式可以适用于固定效应和随机效应(Howell, 2002)。

方差分析还有一些其他可能的效应强度测量方法。这些方法包括同样与 R^2 有关的 Eta 方(η^2),以及与 Cohen d 有关的标准化效应值测量即 ϕ'。

ω^2 是一种比 η^2 偏差更小的效应大小测量方法(参见 Fowler, 1985)。这意味着它是对总体中处理效应的更佳估计,与在当前样本中发现的效应大小正好相反。ω^2 和 η^2 之间存在很多差异,但当我们希望对一个总体推断下结论时,ω^2 总是更小而且更加有用。

——Martin Le Voi

(叶鹏飞译 高勇校)

*也可参见偏倚【Bias】、Eta(η)、固定后果模型【Fixed-Effects Model】、R 方【R-Squared】、方差【Varance】。

参考文献

Fowler, R. L. (1985). Point estimates and confidence intervals in measures of association. *Psychological Bulletin*, *98*, 160-165.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Duxbury, UK: Thomson Learning.

遗漏变量(Omitted Variable)

在一些研究中,一个变量对另一个变量的估计效应会在引入第三个变量后发生改变。当控制第三个变量时,两个变量之间的关系方向可能会发生改变或者这种关系会消失不见。社会科学家必须意识到在研究设计中未包括相关变量的危害。如果未包括相关变量,可能会导致遗漏变量偏倚。

假设我们要估计自变量【Independent Variable】X_1 对因变量【Dependent Variable】Y 的效应。X_1 对 Y 的效应记作 β_1。如果进行回归分析,那么 β_1 的估计值为 b_1。如果研究中存在遗漏变量偏倚,那么估计的 b_1 就是不正确的,因为我们没有考虑另一个自变量 X_2 的影响。

我们怎样才能知道忽略 X_2 是否会导致对 β_1 估计的偏倚?如果 X_2 对因变量 Y 没有影响,那么省略 X_2 就不会引起偏倚。换言之,如果一个自变量与研究的现象不相关,那么排除这个自变量将不会导致偏倚。

只有遗漏变量与解释变量 X_1 相关时,它才会引起对 β_1 的估计偏倚。只要 X_2 与 X_1 相关,在回归方程中引入 X_2 就将改变 β_1 的估计值。如果 X_2 与 X_1 不相关,我们就可以在分析中放心地省略它,即使 X_2 与 Y 之间有很强的关系。因此,可以省略与解释变量不相关的变量,即使这些遗漏变量与因变量之间有较强的关系。只有我们未考虑的变量与解释变量相关时,才会引起遗漏变量偏倚。

然而,如果我们的目标是预测 Y 值,那么我们应当在回归分析中引入每一个与因变量有强烈关系的解释变量。尽可能多地解释因变量的变化,会增加我们预测的准确性,并且如果我们列入了所有相关的解释变量,就会解释更多的变化。当预测【Forecasting】是我们的目标时,应当列入所有与 Y 有较强关系的自变量。如果不想预测 Y 值,或者仅仅想知道 X_1 对 Y 的效应,那么我们可以放心地排除任何与 X_1 不相关的变量。

——Megan L. Shannon

(叶鹏飞译　高勇校)

参考文献

King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research.* Princeton, NJ: Princeton University Press.

单侧检验(One-Sided Test)

参见单尾检验【One-Tailed Test】。

单尾检验(One-Tailed Test)

一个虚无假设【Null Hypothesis】通常用来提出一个特定的总体参数【Parameter】值。我们不一定要相信这就是该参数的真实值,但是如果我们拒绝了虚无假设,并因此排除了这一特定的参数值,那么结果就是这个参数必然等于其他值。在关于两个变量之间关系的研究中,如回归【Regression】当中,虚无假设通常这样表述,即两个变量之间没有关系。用符号表示为 $H_0:\beta=0$,这里 β 是指总体回归线的斜率【Slope】。

如果我们能够拒绝虚无假设,那么就证明 β 不等于 0,意味着变量之间存在关系。接下来,如果总体斜率不等于 0,那么我们如何判断这个斜率的值?一种可能是对这个斜率一无所知,意思是它可能大于 0,也可能小于 0。另一种可能是我们对这个斜率有更多的了解。即根据以前的研究,我们知道这个斜率不可能是负数。然后我们可以把备择假设【Alternative Hypothesis】表述为 $H_1:\beta>0$。因为虚无假设已经被拒绝,所以从备择假设得出的结论是,β 一定大于 0。

这是单尾(也称为单侧)检验的一个例子。之所以称为单尾或单侧,是因为备择假设的单侧形式。备择假设所包括的数值,仅仅是从一个方向上偏离虚无假设中的参数值。

当提出一个单侧备择假设时,只需要一个范围的检验统计量来拒绝虚无假设。当给出一个正态检验统计量和 5% 显著性水平【Significance Level】,配有双尾备择假设的虚无假设在 $z<-1.96$ 或 $z>1.96$ 时要被拒绝。该显著性水平分成两半,分别位于检验统计量的每个尾部,当计算得到大的负值或大的正值时,我们都拒绝虚无假设。但是,对于 5% 显著性水平的单尾检验,拒绝虚无假设的条件是 $z>1.64$。意思是拒绝域仅仅位于检验统计量分布的一尾(正的)。这是因为我们利用了关于这个参数的其他常识,即它是大于 0 的。

对于一个正态检验统计量,双尾备择假设和单尾备择假设之间的差别可能非常重要,比如说,如果 $z=1.85$。在采用双尾备择假设时,虚无假设不会被拒绝,但如果是单尾备择假设,它就要被拒绝。但是这种情况不会经常发生。

另外,随着由预先选择的显著性水平到从数据中计算出来的 p 值【p Value】的转变,单尾和双尾备择假设之间的差别已不那么重要。如果我们报告了一个单尾 p 值,那么读者很容易通过乘以 2 将它转化为双尾 p 值。统计软件一般都会清楚地呈现它计算的是单尾还是双尾 p 值。

——Gudmund R. Iversen

(叶鹏飞译 高勇校)

参考文献

Mohr, L. B. (1990). *Understanding significance testing* (Sage University Papers series on Quantitative Applications in the Social Sciences, series 07-073). Newbury Park, CA: Sage.

Morrison, D. E., & Henkel, R. (Eds.). (1970). *The significance test controversy*. Chicago: Aldine.

单因素方差分析(One-Way ANOVA)

方差分析,简写为 ANOVA,是罗纳德·费希尔(Fisher,1925)首先提出的,用来判断一个或多个因素的均值及因素间的相互作用是否与预期的情况显著不同。"单因素"是指分析中仅包括一个因子。"双因素"(或多因素)是指分析中包含两个(或两个以上)因子以及这些因子间的相互作用。一个因子包含有两个或多个组或层。方差分析基于这一假设,即每个组都来自正态分布的得分总体。此外,不同组得分的方差应当是相等的或齐性的。

不同组的得分可能是相关的或不相关的。相关的得分可能来自相同的状况或者在某些方面相匹配的状况。相关得分的一个例子是,同一变量(比如工作满意度)在两个或多个时间点进行测量。不相关或独立的得分来自不同的情况。不相关得分的一个例子是,同一变量在不同的人群组别中(比如不同职业的人群)进行测量。仅包含两个群组的方差分析等同于假定方差相等的 t 检验。

针对不相关得分的单因素方差分析是将组间方差的总体估计值与组内方差的总体估计值进行比较。这个比值就是用来纪念费希尔的 F 检验、F 统计值或 F 值,其计算公式如下:

$$F = \frac{\text{组间方差估计值}}{\text{组内方差估计值}}$$

组间方差相对于组内方差的估计值越大,F 值就越大,并且越有可能具有统计显著性。F 值的统计显著性是由两组自由度决定的:组间方差估计(F 值的分子或上半部分)的自由度和组内方差估计(F 值的分母或下半部分)的自由度。组间的自由度【Degrees of Freedom】是组数减去 1。组内的自由度是个案数减去组数。

在针对相关得分的单因素方差分析中,F 值的分母是组内方差的总体估计值减去个案产生的方差(the variance due to the cases)总体估计值。这个分母的自由度是用个案数减去 1 乘以组数减去 1。

在因子由两个以上的群组构成并且 F 是统计显著的情况下,哪些均值之间彼此存在显著的差异,需要通过两两比较进行判断。在有充分的根据预测这些差异方向的情况下,这些比较应通过适当的 t 检验来进行。否则,这些比较应通过事后检验(post hoc test)来完成,比如 Scheffé 检验【Scheffé Test】。

——Duncan Cramer

(叶鹏飞译 高勇校)

参考文献

Fisher, R. A. (1925). *Statistical methods for research workers*. London: Oliver and Boyd.

在线研究方法(Online Research Methods)

计算机技术和互联网的发展极大地增强了社会科学家发现信息、与同行合作以及通过计算机辅助的实验和调查收集数据的能力。一些不同类型的工具和技术属于在线研究方法的范畴,包括互联网搜索引擎、在线数据库和数据归档,以及在线实验。虽然互联网本身作为一个网络化计算机的系统集合,已经以某种形式存在了至少 25 年,

但是直到1990年代早期，万维网（World Wide Web）的发展才明显提高了它的可用性，并使相互联通的信息开始爆发式增长。同时，强大的个人计算机的发展，也将开发复杂数据分析的能力带进台式计算机。综合起来，这些技术进步为社会科学家提供了重要的新工具系列。

互联网搜索引擎

互联网本质上是相互联通的计算机网络的集合，这些计算机网络由很多单个的计算机系统组成。将单个计算机系统中的信息放置在服务器上，互联网用户就可以获得相应信息。服务器有几种类型，包括网站（Web）服务器、新闻和文件传输协议（FTP）服务器。从2002年起，大约有200 000 000台服务器（或主机）可以通过互联网访问。为了在庞大的资料集中发现所需材料，大范围的搜索引擎已经被开发出来。搜索引擎通过输入关键字或一个句子，返回一个包含搜索语句的网页列表。某些搜索引擎，比如Google和AltaVista，是首选的；它们使用非常简单的用户界面，并且试图通过某类最优算法返回最匹配的结果。其他引擎，比如和Yahoo！有关的引擎，则更具有综合性，试图通过类别来整理网站，此外也包括关键字和类别搜索。其他知名的搜索引擎还包括Lycos和AskJeeves。和Yahoo！一样，Lycos不仅是搜索引擎，而且提供门户服务，用户不仅可以搜索网站，还可以获取新闻和一系列在线服务。搜索引擎技术的最新发展，包括可以搜寻诸如图片之类的图像资料，还包括搜寻嵌在网站中的字处理文档中的文本。

使用互联网搜索引擎和科学一样是一门艺术，因为对于很多搜索者，可能都会返回数以百计甚至成千上万的潜在匹配信息。一般来说，使用的关键词或短语越明确越好，但是目前搜索引擎技术的状况，使用户不得不经常费力地在大量匹配信息中寻找他们确切需要的部分。事实上，有一个全面的互联网网站，即www.searchenginewatch.com，致力于帮助研究者最有效地使用不同的搜索引擎。

在线数据库/数据归档

互联网的发展极大地提升了科学家共享数据的能力。现在很多数据都保存在万维网上，并且可以通过一般性的或专门化的搜索引擎来获取。例如，全国选举研究（The National Election Studies）提供了自从1948年以来收集的完整数据的自由入口。研究者只要在网站上注册就可以获得使用权。美国联邦政府的综合印刷局（The General Printing Office，GPO）创建了一个单一入口，用来搜索一系列在线数据库，包括国会议事录和联邦公报。这个站点也体现出在线数据库的一个典型问题，即缺乏完整性。例如，GPO联邦公报数据库直到1994年才开始建立，并且其他一些组成数据集的部分甚至最近才开始建立。因此，研究者希望获得的重要历史资料通常不能在线查到。即便如此，如同美国人口普查一样有广泛基础的以及如同维多利亚时代在线数据库一样专门化的数据集，已经从根本上改变了研究者获取信息的方式。

对于学术研究者，很多期刊已经开始提供在线的全文数据库，由该期刊专门维护或者作为多种期刊集的一部分来维护。JSTOR是一个大型全文数据库，纳入了275种期刊的页面图片，共包含73 000期，其中很多期刊可以追溯到它们的第一期。EBSCO则提供了一种更综合性的服务，它提供了大约1 800个期刊的全文入口以及超过1 200种期刊的索引。但是，在JSTOR中可以向前获取很多期刊的第一卷，而

EBSCO 的全文数据仅能回溯到 1985 年,为了弥补这一点,EBSCO 提供现刊的使用权,而 JSTOR 则限制使用至少 3 年前的期刊。大部分大学图书馆可以为用户提供这些数据库服务,同时还包括各种其他服务,包括 Lexis-Nexus(法律研究、新闻期刊和参考信息)与 ERIC(教育)。

在线实验

尽管实验在一些社会科学领域成为一个标准工具已经有很多年,但是将实验在线化后,可以增加参与的被试者数量和多样性,同时也会增加实验设计中的复杂性。在跟踪被试者的响应方式时,计算机可以被用来控制呈现给被试者的文本和非文本材料。利用计算机开展实验简化了数据收集工作,因为实验可以被设计为在实验过程中直接生成数据集,这样就避免了数据录入错误。决策制定和实验经济学的研究者已经开拓性地进行了很多在线实验,也许因为他们的领域支持简单的实验设计以及可以快速完成的实验。即便如此,网络调查【Internet Survey】的数量很可能远远多于真正在互联网上开展的实验数量。汉诺威学院(Hanover College)心理学系的网站上保持有一系列基于互联网的知名心理学项目,其中很多是调查而不是实验。

虽然如此,实验中互联网的使用正在迅速增长。诸如 MediaLab 和网络指南(Web guides)(Kevin O' Neil, *A Guide to Running Surveys and Experiments on the World-Wide Web*)之类的软件,能够简化开展在线实验的过程。然而,愿意使用这一技术的研究者必须考虑几个问题。实验是否适合于在没有实验者的直接个人指导下进行?参与者只能利用网站提供的信息,在实验过程中自己处理问题。实验能否合理快速地完成?目前大部分在线实验时间都相对较短,网络用户通常不会把重要的时间花在这些实验上。实验能否设计成在实验条件下确保随机分配?虽然随机分配【Random Assignment】在任何实验【Experiment】中都是重要的,但是在互联网实验中,被试者的招募无法像传统实验中那样容易进行控制,因此随机分配可能更加至关重要。一个特别引人关注的挑战是从潜在的被试者那里获得知情同意【Informed Consent】。对于为了教学目的而用学生来做的实验,这不是问题,但是对于在研究项目中收集数据用于发表以及将公众作为被试者,必须获得知情同意。已经成为实验前置部分的标准做法是提供知情同意书文本,并要求被试者在屏幕上单击按钮表明同意文件内容。密西西比大学的心理学试验利用这一方法,同时还要求被试者确认单击发送数据的按钮,将他们的结果收入数据集中。

——David P. Redlawsk

(叶鹏飞译　高勇校)

*也可参见计算机辅助个人访谈【Computer-Assisted Personal Interviewing】、网络调查【Internet Survey】。

参考文献

Birnbaum, M. H. (2001). *Introduction to behavioral research on the Internet*. Upper Saddle River, NJ: Prentice Hall.

Krantz, J. H., & Dalal, R. (2000). Validity of Web-based psychological research. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 35-60). New York: Academic Press.

Musch, J., & Reips, U. (2000). A brief history of Web experimenting. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 61-88). San Diego, CA: Academic Press.

Nesbary, D. K. (2000). *Survey research and the World Wide Web*. Boston, MA: Allyn and Bacon.

University of Mississippi. (n.d.). PsychExperiments. Retrieved from http://psychexps.olemiss.edu.

本体论、本体论的(Ontology, Ontological)

本体论是哲学的一个分支,关注的是存在的本质。它是关于存在的理论、什么是实在的理论研究。在社会科学的语境中,所有理论和方法论立场都提出了关于什么样的事物确实或能够存在、它们存在的条件以及它们相互关联方式的假设(隐含的或者明晰的)。

关于社会实在之本质的理论存在两种类型:存在被认为是物质现象的集合或者是人类关于他们世界的观念集合。从人性的视角看,唯物论立场假设自然和社会现象都是独立存在的,并且两类现象都潜在地制约着人类行动(参见唯实论【Realism】)。自然限制的形式是地球引力、气候和我们的身体,而社会限制则包括文化、社会组织和生产体系。唯物论是与自然主义学说联系在一起的,自然主义声称,因为在非生命物的行为和人类行为之间只存在很小的差异,所以自然科学中适用的探索逻辑也能够运用在社会科学中。另一方面,唯心论立场声称在自然现象和社会现象之间存在着根本的不同,人类拥有文化,并且生活在他们共享解释的世界上(参见唯心论【Idealism】)。社会行动不仅是行为,而且是意义传递的过程。正是这些意义构成了社会实在(Johnson, Dandeker, & Ashworth, 1984, pp. 13-15)。

本体论假设是所有社会理论和方法论立场的基础;不同的理论和方法论领域各有不同的本体论主张。例如,实证主义【Positivism】和证伪主义【Falsificationism】蕴含着这样的本体论假设,即有序的世界是由离散的和可观测的事件组成的。人类行动被认为是发生在可观察的物质环境中的可观测行为。社会实在被看成事件之间因果关系的集合体,这里的事件被描述为变量间关系的拼凑物。另一方面,诠释主义【Interpretivism】则假设社会实在是人类就行动和情境的意义进行协商互动这一过程的产物。人类经验的特征是阐释的过程,而不是对外部物质世界的直接感知。因此,社会实在不是一些可以用不同方式进行阐释的"事物",而就是这些阐释本身(Blaikie, 1993, pp.94, 96)。

因为本体论的主张不可避免地与认识论主张(参见认识论【Epistemology】)有关联,所以很难单独讨论它们(Crotty, 1998, p. 10)。关于什么构成社会现象的断言,暗示着存在一定的方法来获取关于这些现象的知识。本体论和认识论主张的差异,无法通过经验的探究加以解决。尽管它们都愿意接受哲学争论,但是不同立场的支持者最终将他们的主张变成为一种信仰行动。

——Norman Blaikie

(叶鹏飞译 高勇校)

参考文献

Blaikie, N. (1993). *Approaches to social enquiry*. Cambridge, UK: Polity Press.

Crotty, M. (1998). *The foundations of social research*. London: Sage.

Johnson, T., Dandeker, C., & Ashworth, C. (1984). *The structure of social theory*. London: Macmillan.

开放性问题(Open Question)

参见开放性问题【Open-Ended Question】。

开放式问题(Open-Ended Question)

开放式问题也称作开放性、无结构式或定性的问题,是指那些响应类型或回答类别由受访者而非访谈者提供的问题。与它相对照的是封闭式或结构式问题,由访谈者提供有限的响应类别,受访者从中作出选择。因此,受访者可以用他们自己的措辞或方式来反映自己观念的方式,而不是研究者的术语和方式来为开放性问题提供答案。

这类问题在面对面的访谈中非常有效,但在邮件、网络或电话访谈中效果不佳。当研究是以现象学的目标为导向时,开放式问题特别有效。人们经常喜欢发表自己的观点,设置开放式问题对于受访者能产生潜在的满足。此外,开放式问题常常会产生意想不到的叙述或响应类别。而且开放式问题在以下几种情况中是非常合适的,即当可能的回答范围超出研究者提供的响应列表(例如娱乐偏好)时;当问题需要一个详细的、扩展的响应时;当研究者希望发现受访者对某个话题的知晓情况时;以及当有必要充实某个复杂问题的细节时(Fowler, 1995)。开放式问题还在产生封闭式或结构式问题的回答类别时非常有用。

不管怎样,开放式问题对受访者的要求更高,特别是对那些受教育不多的人(Frey, 1989)。回答的细节和深度也会受到受访者对研究课题的熟悉程度或投入程度的影响。开放式问题会产生冗长的、详尽的回答,其中很多内容可能与主题无关,也可能很难进行编码。访谈中的开放式问题通常需要花费更长的时间去询问和记录回答,因此会增加调查的成本。成本可能是开放式问题在当前的调查研究中被有限使用的主要原因;当开放式问题是获取某些回答的唯一方法时,调查研究者会采用这种形式。

最后,开放式问题逐渐成为一种"语言事件(linguistic event)",在这里,访谈者不是简单地作为联结研究问题的中立或被动的渠道。更确切地说,访谈成为一种反映受访者与访谈者互动的协商式文本(Fontana & Frey, 2000)。访谈是一种经由开放式问题而得到扩展的社会交往,这一过程有可能提供一种不代表受访者关于事件看法的结果;在某种意义上,它被访谈者与受访者的互动所扭曲。

抛开那些与费用、成本和情境效应相关联的难题,开放式问题是很有价值的信息获取方式。有些困难可以通过对响应增加一些结构或限制(Fowler, 1995),以及在开放式问题的使用上对访谈者进行适当训练而加以克服。

——James H. Frey

(叶鹏飞译　高勇校)

参考文献

Fontana, A., & Frey, J. H. (2000). The interview: From structured questions to negotiated text. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 645-672). Thousand Oaks, CA: Sage.

Fowler, F. J. (1995). *Improving survey questions*. Thousand Oaks, CA: Sage.

Frey, J. F. (1989). *Survey research by telephone*. Newbury Park, CA: Sage.

操作定义(Operational Definition)

参见概念化【Conceptualization】、操作化【Operationalization】和测量【Measurement】。

操作化论(Operationism/Operationalism)

操作化论诞生于自然科学中珀西·布里奇曼的研究工作,是实证主义的一个变种。它指明科学概念必须链接到工具性的程序,以便确定它们的取值。"虽然诸如'质量'这样的概念可以在理论上或形而上学地表达为一种性质,但这只是人们虔诚相信的观点……作为性质的'质量'等于从指针读数中推断出来的'质量'"(Blalock,1961, p.6)。

在社会科学中,通过社会学家乔治·伦德贝格(Lundberg,1939)的工作,操作化论受到短暂的欢迎。他断言社会学家的下述观点是错误的,即认为测量只能在事物被适当定义之后才能实施——例如,在对异化进行测量之前必须要有一个可操作的定义。对于伦德贝格来说,定义是通过测量产生出来的。他认为社会价值观可以通过考查人们对特定事物的出价来测量(也据此给出定义)。人们看重 *X* 吗?如果是,他们对 *X* 的出价是多少?

操作化论在自然科学和社会科学被实证主义【Positivism】主导时基本上是没有争议的,但它随着实证主义的失势也已经不再风光。在社会科学中,已经确认存在三个主要困难。第一个是不完全决定性(underdetermination),自然科学中也存在这一问题。研究者如何能确信可检验的命题已经充分地将一种理论操作化了?诸如无家可归、贫困和异化这样的概念,在不同的社会环境中具有一系列不同的维度和表现。以无家可归为例,它要取决于历史上和社会中对"收容所"与"家"概念之间的变化关系的定义。第二个问题是现象的客观外部定义可能不会(甚至是无法)与那些经过体验给出的主观定义相一致。对一个人来说看似是贫困的状况,对另一个人可能就是富有。第三个问题是社会科学家之间对定义的分歧。尽管个别研究者可能会找到第二个问题的局部实用性的解决方案,但是他们的方法很可能与其他研究者的方法存在差异或者相互矛盾。

这些困难似乎对这一方法论提出了不可否认的反驳,尽管操作化论已经日渐式微,但它仍然是一个未解决的问题。所有调查研究者都熟悉操作化的过程,即从理论到测量"降低抽象的层次"。这个过程逻辑上等同于操作化论的过程。例如,一个综合性的、非操作化的无家可归理论可以用陈述的方式进行表达,但如果一个人希望利用一项社会调查来测量或描述无家可归,那么必须对无家可归进行操作性的定义(Williams,1998)。操作化和操作化论的区别可能存在于理论与研究的关系中。操作化论认为理论构建与测量之间应当有劳动分工,而大多数调查研究者强调理论建构与检验理论的变量设定之间应当有紧密的动态关系。

——Malcolm Williams

(叶鹏飞译　高勇校)

参考文献

Blalock, H. (1961). *Causal inference in nonexperimental research*. Chapel Hill: University of North Carolina Press.

Lundberg, G. (1939). *Foundations of sociology*. New York: Macmillan.

Williams, M. (1998). The social world as knowable. In T. May & M. (Eds.), *Knowing the social world* (pp. 5-21). Buckingham, UK: Open University Press.

最优匹配(Optimal Matching)

最优匹配(有时被称为最优比对)是用来分析两个数据序列之间整体相似性的一种计量技术。该方法使得长度不均匀序列或包含重复数据序列的比较成为可能。与尺度或聚类算法【Algorithm】相结合,最优匹配能够在很多情况下用于发现一般序列模式。

最优匹配方法是由弗拉迪米尔·莱文斯坦在1965年首先提出的,此后被广泛应用于自然科学中DNA序列的相似性识别,在计算机科学中的应用是作为词语和语言识别算法的基础。安德鲁·阿博特(Abbott, 1986)将这一方法引入社会科学中。

最优匹配依赖的前提是,序列数据中的相似性能够通过计算一个序列转换成另一个序列的容易程度来进行量化。一个序列向另一个序列的转化,是通过一系列基本运算来完成的,通常包括插入元素(I_i)、删除元素(D_i)以及代替另一个元素的替代元素(S_{ij})。这些运算的每一次使用都涉及一个固定成本(对于每个元素 i 或元素对 ij 可能是不同的)。利用与这些运算相关的成本,最优匹配算法可以确定将一个序列转化为另一序列的最小总体成本;这个值通常被称为这对序列之间的距离。距离小的序列对彼此间的相似性高于距离大的序列对。

如果我们将单词看成字母序列,就可以运用最优匹配来计算PUTTER和PATTERN之间的相似性(图1)。如果每次运算要求相同的成本(对于所有的元素 i 和 j,都有 $S_{ij}=I_i=D_i$),很容易看到,在这个例子中,最简单的比对包含两次转换:在PUTTER的第2个位置,将U换成A,在其第8个位置插入N。

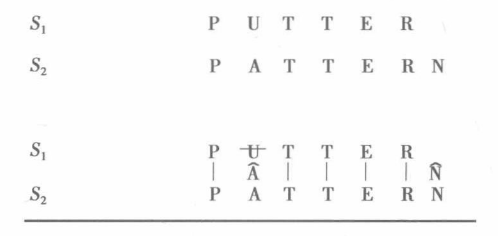


图1 最优匹配

尽管在某些情况下,给每次转换分配相同的成本是合理的,但大部分社会科学中的应用都要求更复杂的成本设计。例如,最优匹配常常被用于揭示职业结构的模式,其中职业是由工作来构建的。从经验上看,一些工作比其他工作更为流行——也因此更容易被"替代",一个通用的解决办法是把元素(工作)的总体排列成线性的尺度,然后通过计算它们在各尺度上的分值差异,来生成每对元素的替代成本矩阵【Matrix】。于是 $S_{ij}=|i-j|$。在其他诸如集体行动事件顺序或政策变迁历史的大量情境中,增加或减少边际成本的非线性替代方案可能是更为适当的。一般地,插入和删除成本是相等的,并且被设置成接近最大替代成本的值。

尽管最优匹配基本上是一种二值技术(dyadic technique),但它能够被应用于含有很多序列数据个案的大型数据集。在这种情况下,最优匹配算法会产生一个矩阵,其中每个单元都是由一对序列之间的距离所组成。正如任何距离矩阵一样,尺度或聚类算法都可以用来检测成群的个案之间彼此接近程度;需要说明的是,个案之间的距离越小,反映的是数据中基础序列的相似性更强。也可以根据个案与理论推断的序列或经验性的一般序列的相似性,对这些个案进行分类;这种分类可以在标准回归模型中作为变量使用。

——Katherine Stovel
(叶鹏飞译　高勇校)

参考文献

Abbott, A., & Forrest, J. (1986). Optimal matching methods for historical sequences. *Journal of Interdisciplinary History*, *16*, 471-494.

Abbott, A., & Hrycak, A. (1990). Measuring resemblance in sequence data. *American Journal of Sociology*, *96*, 144-185.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Cybernetics and Control Theory*, *10*, 707-710. (Original work published 1965.)

Sankoff, D., & Kruskal, J. B. (1983). *Time warps, string edits, and macromolecules*. Reading, MA: Addison-Wesley.

最优度量(Optimal Scaling)

最优度量是达雷尔·博客在 1960 年为度量多元定类数据而创造的术语,在数学上相当于双重度量【Dual Scaling】、对应分析【Correspondence Analysis】和发生率数据的同质分析。

——Shizuhiko Nishisato
(叶鹏飞译　高勇校)

口述史(Oral History)

口述史是通过记录被访者关于过去的记忆来研究历史,以描绘关于事件和情境的直接体验和个人经历。它在历史学、社会学、文化研究和心理学的交叉领域占有一席之地。支持口述史者发起了挑战,认为关于过去的官方解释和主流解释都是基于历史档案的。对于许多研究者来说,这一方法的魅力在于其解放性和批判性的主旨(Perks & Thomson, 1998; Thompson, 2000)。批评者则质疑这些建立在幸存者生平记忆基础上的资料的效度。

访谈及其分析

口述史的核心是访谈【Interviewing】,访谈不仅是社会科学实践中一种主要的资料收集方法,同时也为研究当代史和 20 世纪历史的史学家们所熟悉。对于口述史家来说,访谈的本质特征是它的即时表达、社会关系和天然的询问性质,而且口述史具有广泛的学科包容性。根据口述史家的不同学科基础和选择的主题,寻找访谈对象以及进行访谈的方式可能多种多样。

在一项早期研究中,保罗·汤普森(Thompson,2000, pp.145ff)使用了一个配额样本【Quota Sample】,他采用社会学的方法从一个无法进行随机选择的总体中找到了受访对象。当口述史家记述幸存者晚年记忆时,他们是在相互合作。在汤普森的实例中,配额样本是按照预先设定的访谈对象类别来确定的,该类别与1911年英国人口普查认定的职业类别是一致的。随着时间的推移,当研究焦点集中于一般社会趋势时,这样的方法备受人们喜爱,如对家庭关系、性别、移民、就业或政治世代的研究。

在类似研究中,访谈倾向于根据标准化的提问进行,话题领域已经确定,并且经常是在生活史【Life History】的视角范围内。访谈一般要遵循严格的规定格式,因为它的目的是提炼出具有可比性的证据材料,在呈现出来的现象中寻找规律和模式。这种类型的研究可能需要100个左右的样本量作为基础。

与此相反,有些研究则倾向于使用一种完全不同的方法来寻找和访谈对象,比如寻求对特殊事件的探析,记述一群精心挑选出来的幸存者经历,或关注某一个体的独特故事。在这样的案例中,寻找被访者最常用的方法是滚雪球抽样【Snowball Sampling】,即通过媒体求助和使用会员名册、代表机构或老年人组织的网络来招募被访者。

当研究仅仅针对很少几个对象时,访谈的方法可能截然不同。首先,可能需要对同一人进行不止一次的回访,目的是获取一个完整的故事,并且使被访者能够自由表达自己的主观经验。这样,方法就需要相应改变。其次,对某些口述史家来说,一种趋向心理分析的、自由联想的互动更具吸引力。挖掘类似童年分离、基于种族和残疾的隔离经历以及罢工或冲突这类事件的记忆,可以释放强烈的情感,特别是当它们与其他证据、文献或视频材料关联起来时,或者在访谈中将它们定位和描述为整个人生经历的一部分时。

对沉默和压抑体验的鉴别,引起了口述史家的兴趣,无论他们是否受到心理分析理论的影响。口述史家中间普遍出现一种转变,开始接受更加主观有时甚至是虚构的故事(Perks & Thomson, 1998)。认可这类资料,承认这类资料既有文化的重要性也有心理学的重要性,代表着一种对经验和实证方法的抛弃,进而又影响对资料的解读和聚焦于效度的批判。

对访谈资料的分析通常采用主题法,这种方法被历史学家和社会学家用于确定主题——根据研究设计或选择的理论框架预先确定,或者基于扎根理论【Grounded Theory】从资料本身推演而出(Thompson, 2000, p. 151)。尽管这里再次强调语言和表达的重要性,但是计算机方法的使用变得越来越普遍,特别是面对大样本的访谈。研究者将他们独有的学科基础——例如历史学、心理学、社会学或社会政策——所提供的解释力,与其他资源相互参照,包括文献的或口头的,以求从历史和结构的视角来界定和验证他们的分析。

进展和争议

与传记法【Biographical Method】、叙事分析【Narrative Analysis】、人类学和民族志【Ethnography】的进展一样,口述史研究者已经展开了有关研究方法的讨论,尤其是在女权主义研究【Feminist Research】、伦理学和所有权的领域。

女性作为研究对象以及作为研究者的身份,引发了对访谈过程中的影响力的争论。本质主义的立场认为,女性可以与她的访谈对象共享理解和经验,作为研究者的身

份能够引起特殊的共鸣。一个相反的立场——和女性主义者一样,尽管更具批判性和反思性——认识到访谈关系中存在显著差异的其他方面,尤其是阶级、肤色和世代。肖娜·伯吉尔·格拉克和达芙妮·保陶伊(Gluck & Patai,1991)编辑的论文集对促成这些辩论作出了重要贡献,其中包括一些从事口述史研究的女性主义者深度自我批判的内容。

从许可和所有权问题中衍生出来的伦理议题,目前在口述史研究中扮演着一个核心角色。按照欧盟法律,访谈中的资料版权属于被访者,这意味着除非签署协议,明确版权属于访问者或者其所在的工作机构,否则不能公布任何有关访谈的资料(Ward,2002)。这一要求是某些口述史家所认为的核心伦理议题的法律表述,即被访者要同意参与访谈和记录资料的使用。让被访者能够掌控访谈录音和文字记录的归宿,被认为是访谈过程中的一个必要部分,相关的标准文本可以在口述史网站上找到,与档案【Archive】存放和调取有关的规章文本中也可以找到。

——Joanna Bornat

(叶鹏飞译　高勇校)

参考文献

Gluck, S. B., & Patai, D. (Eds.). (1991). *Women's words: The feminist practice of oral history*. London: Routledge.

Perks, R., & Thomson, A. (Eds.). (1998). *The oral history reader*. London: Routledge.

Thompson, P. (2000). *The voice of the past* (3rd ed.). Oxford, UK: Oxford University Press.

Ward, A. (2002). *Oral history and copyright*. Retrieved June 7, 2002 from http://www.oralhistory.org.uk

阶(Order)

阶是指在两个变量关系的统计分析中被控制或保持不变的附加变量的数目。零阶关系测量两个变量之间关联的大小或强度,不去控制任何其他因子(Knoke, Bohrnstedt, & Mee, 2002, p. 213)。常见的例子包括两个分类变量的列联表【Contingency Table】、两个连续变量之间的相关【Correlation】和一个自变量对一个因变量的简单回归【Regression】。

一阶关系涉及控制第三个变量的影响来判断一个二变量关系是否发生变化。在交叉表分析中,这个过程被称为详析【Elaboration】(Kendall & Lazarsfeld, 1950)。变量 X 和 Y 之间关系的详析是通过将它们的零阶交叉表分成一阶子表来进行的,也称为偏表(partial table)。每个子表都表示在一种类型的控制变量 Z 之下的 X-Y 关系。如果 X-Y 的关系不受 Z 的影响,那么这种关系的测量,诸如 Phi 系数【Phi Coefficient】或 Gamma 系数【Gamma Coefficient】,就会在每个一阶表中有相同的值。如果由于 Z 对两个变量的影响导致这种关系是虚假的,那么它就会消失在每个子表中。因此,Z“解释”了 X-Y 的关系。如果 Z 干预了 X 到 Y 的因果顺序,就会出现相同结果;因此,Z“阐释”了 X-Y 的关系。详析通常揭示一种“特定性(specification)”或交互效应【Interaction Effect】,其中关系的强度或方向在子表间是

不同的,说明 X-Y 的关系根据 Z 的特定类别而具有偶然性和条件性。

三个连续变量的一阶关系可以通过偏相关【Partial Correlation】系数进行估计,偏相关系数表示的是"在一个控制变量的影响被保持不变的情况下,两个变量之间共变关系的大小和方向"(Knoke et al., 2002, p. 223)。其计算公式是三个零阶相关的函数:

$$r_{XY \times Z} = \frac{r_{XY} - r_{XZ} r_{YZ}}{\sqrt{1 - r_{XZ}^2}\sqrt{1 - r_{YZ}^2}}$$

如果 X 和 Y 都与变量 Z 不相关,那么它们的偏相关等于其零阶相关,$r_{XY \times Z} = r_{XY}$;也就是说,它们的线性关系不受该控制变量的影响。但是,如果 Z 与两个变量都存在相同方向上的强相关,那么它们之间减弱的(或 0)偏相关表明 X-Y 的关系部分或者全部是虚假的。

二阶和更高阶关系将这些方法一般化,通过使用多种控制变量用来分析零阶关系。对数线性模型【Log-Linear Modeling】被用来估计多元交叉表的复杂交互效应。在多元回归分析【Multiple Regression Analysis】方程中,各个偏回归系数表示在控制("排除")其他预测变量的叠加效应后,一个自变量发生一个单位的变化对因变量所产生的影响。检测交互效应涉及建立乘积项或其他自变量组合(Jaccard, Turrisi, & Wan, 1990)。

阶还涉及随机变量 X 分布的矩【Moment】。原点矩是根据作为 X^k 期望值的第 k 个正整数或 $E(X^k)$ 来定义的;因此,一阶原点矩 $E(X)$ 就是均值【Mean】。如果在求幂之前用 X 减去均值,那么中心矩就被定义为 $E[X-E(X)]^k$。X 的一阶中心矩等于 0,而它的二阶中心矩就是方差【Variance】。三阶和四阶中心矩分别被称为偏态和峰态【Kurtosis】。

——David H. Knoke

(叶鹏飞译 高勇校)

参考文献

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression.* Newbury Park, CA: Sage.

Kendall, P. L., & Lazarsfeld, P. F. (1950). Problems of survey analysis. In R. K. Merton & P. Lazarsfeld (Eds.), *Continuities in social research* (pp.133-196). New York: Free Press.

Knoke, D., Bohrnstedt, G. W., & Mee, A. P. (2002). *Statistics for social data analysis* (4th ed.). Itasca, IL: Peacock.

次序效应(Order Effects)

在抽样调查【Survey】访谈和问卷【Questionnaire】中,问题和答案次序的效应已成为社会学家、政治学家、民意研究人员和认知社会心理学家的众多方法论实验所关注的焦点。正如在口语和书面沟通中词或短语的意义不能脱离它所出现的特定语境,研究人员发现,在访谈【Interviewing】或有自填式问卷【Self-Administered Questionnaire】时,调查问题的意义会非常显著地受到被访者对前面问题的应对的影响。他们还认识到,因为对答案系列的陈述有不同方式,所以次序效应不仅会出现在调查的问题之间,也会出现在问题之内。受访者显然会根据问题和答案类目的次序来推断问题的

意义。

当被访者在同一主题或者相近话题上被问到两个或更多问题时,问题的次序和语境效应通常就会出现。即使访谈或问卷中的多个题目分隔了那些密切相关的问题,这种效应仍然会出现。最初尝试对这类效应的理解,包括对问题之间关系类型进行概念分类的发展:(a)部分-整体关系,涉及两个或更多问题,其中有一个更具综合性,因此暗含或者包括了其他的问题,但反过来却不一定(例如,某人整体的生活幸福程度与某人的婚姻幸福)。(b)部分-部分关系,涉及在某个特征上处于同一层次的多个问题(例如,允许堕胎以终止因强奸导致的妊娠与终止先天缺陷的妊娠)。相应地确定了两种类型的效应:(a)一致性效应,被访者对后续问题的回答更多地表现为与前面问题的答案相一致(例如,投票偏好和他们的政党认同);(b)差异化效应,他们对后面问题的回答会区别于前面问题的答案(例如,他们赞成乔治·布什处理经济问题的方式与是否赞成他处理外交关系的方式)。最后结果是有四种类型的问题次序效应:部分-部分一致性效应、部分-整体一致性效应、部分-部分对比效应和部分-整体对比效应(Schuman & Presser, 1981, Chap. 2)。史密斯(Smith, 1992)后来论证了次序效应不仅依赖于前面问题的整体语境——正如舒曼和普雷瑟的类型学所确认的,而且依赖于被访者实际上如何回答了前置的问题。从概念上讲,这种问题次序和对前置问题的回答之间的交互作用,被称为"条件性次序效应"。更近期的心理学研究已经在运用认知可及性的启发式方法、信赖抽样(belief sampling)模式、会话逻辑以及同化和对比效应的认知模型,来对各种不同类型的问题次序和语境效应进行理论探讨(如 Sudman, Bradburn, & Schwarz, 1996)。

调查问题中的答案次序效应可能以几种方式出现。当一个问题的多个答案类别被大声宣读时,比如在做电话访谈或者面对面访谈没有使用可见的答案卡片,被访者将倾向于选择最后提到的答案选项:近因效应。如果答案类别以可视化的形式呈现,比如在做一个自填式邮寄调查、网络民意调查或投票后民意调查时,被访者会倾向于选择第一个选项:首因效应。当某个问卷题目前面的另一个题目在某些回答的测量维度上产生了很极端的评价,比如当一个非常喜欢或不喜欢的政治人物出现在一系列被评价的其他人物中间,答案类别间的判断对比效应同样可能出现。答案次序效应的解释,从传统的结晶化假设(the crystallization hypothesis)——认为这些效应在那些对于相关话题事先持有坚定观点的被访者中不太会出现,到记忆限制假设——主张这些效应的出现是因为被访者难以记住所有的答案选项,再到最近的认知精细化(cognitive elaboration)和满意性模型(satisficing models)(Bishop & Smith, 2001)。源于认知社会心理学中关于态度改变研究的认知精细化模型,认为之所以存在回答次序效应,是因为被访者利用这一机会来思考答案选项。以电话调查【Telephone Survey】为例,当一系列答案选项在向被访者大声宣读后,被访者更有可能思考后面听到的选项,这常常会产生近因效应。与此相反,源于西蒙的经济决策理论的满意性模型,假定大多数被访者会选择答案中第一个满意的或者可接受的选项作为回答,而不是花时间去选择一个最佳答案。例如,在电话调查问题中选择最后提到的选项,证明了被访者倾向于满意策略而不是最佳化。尽管出现了这些充满前景的认知理论发展,现有的解释和预测仍

然难以完全解释回答次序效应。

——George F. Bishop

（叶鹏飞译　高勇校）

参考文献

Bishop, G., & Smith, A. (2001). Response-order effects and the early Gallup split-ballots. *Public Opinion Quarterly*, *65*, 479-505.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys*. New York: Academic Press.

Smith, T. W. (1992). Thoughts on the nature of context effects. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp.163-184). New York: Springer-Verlag.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.

定序交互(Ordinal Interaction)

定序交互和反序交互之间的区别经常被关注。这个区别通常出现在回归【Regression】分析的情境中，当要比较一个组的 Y 在 X 上的斜率和另一个组的 Y 在 X 上的斜率时。如果这两个斜率不相等，就意味着存在交互【Interaction】。对于每一组，都可以画出一条回归线来表示 Y 和 X 之间的关系，涵盖所有在组中观察到的 X 值。两组的回归线不平行，就意味着出现交互。反序交互是指其中一组的回归线与另一组相应的回归线发生交叉。这也被称为交叉交互。定序交互是指两组回归线并不平行，但在研究的数据范围内并未出现交叉。

对于任何给定的一对非平行回归线，总会存在一个两线相交的点。在这个意义上，所有交互理论都是非定序的。只有在被研究的数值范围内(例如，在 90 到 110 的 IQ 值)，回归线不会交叉，这些交互才能被归类为定序交互。

如果已知两条回归线的截距【Intercept】和斜率【Slope】值，则可以用代数方法计算出两条回归线交叉处的 X 值。设 $\beta_{0,A}$ 为组 A 的截距，$\beta_{1,A}$ 是组 A 的斜率，$\beta_{0,B}$ 为组 B 的截距，$\beta_{1,B}$ 为组 B 的斜率。则交叉的点是

$$PI = \frac{\beta_{0,A} - \beta_{0,B}}{\beta_{1,B} - \beta_{1,A}}$$

应用性的研究者表达了一些关于定序交互的谨慎看法。他们主张，这些交互可能是结果变量的度量单位造成的假象。通过对结果变量数值的简单单调【Monotonic】转换，非平行回归线经常可以变得平行。如果对结果变量的度量是任意的，那么一些科学家认为，为了模型【Model】的简约性，必须考虑通过转换来消除交互。然而，如果变量的潜在度量是有意义的，那么定序交互不应被摒弃。

——James J. Jaccard

（叶鹏飞译　高勇校）

定序测量(Ordinal Measure)

一次测量是一个数据点。在很多研究应用中，测量是对某个个体、生物体或研究

参与者分配一个数字，反映他、她或它在某种概念上的状况。一次测量是在一个特定时间、对特定个体、在特定环境、使用特定类型的记录装置来进行的。一组测量涉及多个数据点。例如，对 10 个人中每个人的身高进行测量，表示一组 10 个测量。

定序测量可以从不同的角度进行描述。当应用于一组测量时，序（ordinality）描述的是把概念的观察测量值与其真实值联系起来的函数性质。如果某个概念的测量值是其真实值的严格单调【Monotonic】函数，也就是说，当使用观察测量来排序个体时，真实概念中个体的相对顺序也得到保持，那么该概念的一组测量具有序的性质。例如，10 个人的身高可能是不同的，他们可以从低到高进行排序。研究者可能会提出一种测量身高的策略，让指标中更高的得分表示更高的身高。这组对 10 个人进行的测量，如果个体在指标上的排序与个体身高基本维度上的顺序相一致，就可以说具有定序性质。

严格地说，定序测量并不表达个体或生物体之间在基本维度上的差异之数量信息。定序测量仅仅允许我们来表示一个个体在该维度上多于或少于另一个个体。

——James J. Jaccard

（叶鹏飞译　高勇校）

参考文献

Anderson, N. H. (1981). *Methods of information integration theory*. New York: Academic Press.

最小二乘（Ordinary Least Squares, OLS）

考虑一个线性关系，其中，取值为 y_i 的随机因变量，是一个或多个非随机自变量 $x_{1i}, x_{2i}, \cdots, x_{ki}$ 的函数。我们可以将其写作 $y_i=\beta_0+\beta_1 x_{1i}+\cdots+\beta_k x_{ki}+\varepsilon_i$。对于任何这种关系，都有一个预测方程产生关于因变量的预测值。例如，$\hat{y}_i=\hat{\beta}_0+\hat{\beta}_1 x_{1i}+\cdots+\hat{\beta}_k x_{ki}$。预测值通常被标记为符合 $\hat{y}_i$。如果用真实值减去预测值（$y_i-\hat{y}_i$），那么得到一组残差【Residual】（ε_i），每个残差测量 y_i 的一个预测值和相应的真实观察值之间的距离。OLS 是参数点估计的一种方法，它将由关于参数的残差（或距离）平方和所定义的函数最小化。

OLS 估计

例如，考虑简单双变量回归 $y_i=\beta_0+\beta_1 x_i+\varepsilon_i$。预测方程是 $\hat{y}_i=\hat{\beta}_0+\hat{\beta}_1 x_i$。残差计算方程式为 $\varepsilon_i=y_i-\hat{\beta}_0-\hat{\beta}_1 x_i=y_i-\hat{y}_i$。于是，OLS 估计就是通过发现关于参数 β_0 和 β_1 的函数 $S=\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i)^2$ 的最小值来定义。这是一个凹二次函数，总会存在一个单个最小值。这个简单的最小化问题可以得到答案。最小化这一函数的标准方法，是先求两个参数 β_0 和 β_1 的偏导数，然后解出两个二元联立方程的结果集。具体步骤如下：

$$\frac{\partial\left[\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i)^2\right]}{\partial\beta_0}$$

$$=\sum_{i=1}^{n}2(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i)(-1)=0$$

$$\frac{\partial\left[\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i)^2\right]}{\partial\beta_1}$$

$$=\sum_{i=1}^{n}2(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i)(-x_i)=0$$

简化结果并重新排列,得到所谓的标准方程式:

$$\sum_{i=1}^{n} y_i = n\hat{\beta}_0 + \left(\sum_{i=1}^{n} x_i\right)\hat{\beta}_1,$$

$$\sum_{i=1}^{n} x_i y_i = \left(\sum_{i=1}^{n} x_i\right)\hat{\beta}_0 + \left(\sum_{i=1}^{n} x_i^2\right)\hat{\beta}_1$$

因为在前述方程中很容易计算总和,我们只要解出两个二元联立方程,就能得出对回归参数的 OLS 估计。这些分析步骤很容易扩展到多变量的案例中,对应产生一组针对问题维度的标准方程式。例如,参见 Greene(2003, p. 21)。

OLS 估计的经典假设

OLS 估计量事实上已经被整合进每一个统计软件包,并且非常流行。其广受欢迎的原因在于它易于使用且具有非常实用的统计属性。参数估计关注的是通过样本统计量来获得总体的参数值。我们永远无法肯定某个样本统计量一定就代表了总体参数。然而,通过大家熟知的统计原理,可以知道用来产生样本统计量的估计量属性。

根据高斯-马尔科夫定理【Gauss-Markov Theorem】,一个 OLS 回归估计量要成为最佳线性无偏估计量【Best Linear Unbiased Estimator, BLUE】,需要满足以下假设:

(1)总体干扰的期望值是零($E[\varepsilon_i]=0$)。该假设意味着从总体回归函数中省略任意一个变量,不会对干扰的均值产生系统性效应。

(2)总体自变量和干扰之间协方差的期望值为零($E[X_i,\varepsilon_i]=0$)。该假设意味着自变量是非随机的,或者如果它们是随机的,则它们不能与干扰发生系统性的共变。

(3)干扰之间总体相关的期望值为零($E[\varepsilon_i,\varepsilon_j]=0$)。这是所谓的非自相关假设。

(4)干扰的总体方差期望值为常数($E[\varepsilon_i^2]=0$)。这是所谓的同方差(nonheteroskedasticity)假设。

这四条关于总体干扰的假设通常被称为高斯-马尔科夫假设。然而,其他条件也与 OLS 的统计属性相关。如果干扰是正态的,那么 OLS 估计量是最佳无偏估计量(BUE);但如果干扰是非正态的,那么 OLS 估计量仅仅是最佳线性无偏估计量(BLUE)。换言之,对于非正态的特征,另一种非线性或者稳健的估计量可能更为有效。

此外,总体回归函数必须通过样本回归函数准确地加以确定。所有相关的自变量都应包含,但不应包括不相关的自变量。总体回归函数的函数形式应该是线性的。

OLS 估计量的另一个要求是不存在完全多重共线性【Multicollinearity】。完全多重共线性出现在某个自变量是其他一个或多个自变量的完全线性组合时。如果存在完全多重共线性,则 OLS 估计量无法确定,并且样本分布的标准误差无穷大。如果研究者要分解回归量的独立效应,高多重共线性可能是有问题的。然而,即使面对高多重共线性,OLS 估计量也是 BLUE。

运用 OLS 估计量的统计推论

要知道,OLS 估计量是 BLUE 的特点,提供了绝对的确定性,即平均来说,在重复抽样下,估计的样本统计量将会等于总体参数。此外,重复抽样下运用 OLS 得出的样本统计量的方差,在这个方差所有可能的估计量中,将会是最小的。

尽管我们永远不能确定任一特定样本的样本统计量能否代表总体,但是关于产生样本统计量的估计量的知识,使我们能够对估计的置信度进行评估。例如,回归斜率系数的 OLS 抽样分布的标准偏差表示为

$$SE(\hat{\beta}_k) = \sqrt{\frac{\hat{\sigma}_e^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2(1 - R_j^2)}}$$

其中 R_j^2 是变量 j 在所有其他自变量上的回归所解释的方差。这也被称为标准误。运用标准误,我们可以作出关于一个样本统计量代表总体参数的可能性的概率陈述。如果重复抽样下样本统计量的分布是正态的,一个估计偏离真实参数超过标准偏差的概率是很容易计算出来的。然而,在上述方程中分子表示的干扰方差是一个估计而不是真实总体方差的情况下,就应该使用 t 分布。

尽管从技术上来讲 OLS 估计量并不要求误差项是正态分布的,但通常都假定总体干扰的分布是正态的。然而,非正态干扰的结果是相当严重的。"粗尾(fat-tailed)"分布类的非正态性,意味着假设检验失去意义。此时有两种选择:第一,这个数据通常可以通过转换产生正态分布的干扰;第二,可以采用一种可获得的稳健估计量来应对非正态分布的干扰。

——B. Dan Wood
Sung Ho Park
(叶鹏飞译　高勇校)

参考文献

Greene, W. H. (2003). *Econometric analysis* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Gujarati, D. N. (1995). *Basic econometrics* (3rd ed.). New York: McGraw-Hill.

Judge, G. G., Hill, R. C., Griffiths, W. E., Lutkepohl, H., & Lee, T. (1988). *Introduction to the theory and practice of econometrics* (2nd ed.). New York: John Wiley & Sons.

Schroeder, L. D., Sjoquist, D. L., & Stephan, P. E. (1986). *Understanding regression analysis: An introductory guide.* Beverly Hills, CA: Sage.

组织民族志(Organizational Ethnography)

从 19 世纪开始,民族志【Ethnography】("民族描述")就意味着对一个民族的描述:他们的风俗、礼仪和信仰。组织民族志于 1980 年代伴随着对文化、象征和民俗的兴趣而出现。通过运用田野研究【Field Research】,民族志学者通过记录一个组织的传统和象征行为,以提示组织中人的情感、观点、做事的习惯方式及整个组织的精神气质。目的是理解组织行为并促进组织发展。

1980 年代初,出现对思考组织的新范式的探寻:视角上指向人类维度,方法上放弃机械的和超机体的模型。1980 年,《组织管理协会评论》(*Academy of Management Review*)发表了丹德里奇、米特罗夫和乔伊斯的一篇关于"组织象征"的文章。1983 年《管理科学季刊》(*Administrative Science Quarterly*)出版了一期以"组织文化"为主题的特刊。同年,米歇尔·欧文·琼举办了一场研讨会,会议主题是"神话、象征和民俗:组织分析的扩展",介绍了"组织民俗"的概念。尽管约翰·范梅南(Van Maanen,1979)更早使用了"组织民族志"这一术语,但位于华盛顿州奥林匹亚的常青州立学院(Evergreen State College)应当受到赞誉,它是第一个公开对此进行实践的机构。1990 年 5 月,该机构寻求招聘一名民族志学者来探寻自身的文化,最终选择了民俗学和神话学的博士候选人彼得·特默拉普(Tommerup,1993)。通过围绕故事、仪式和其他表现形式进行访谈和观察,他收集到可以描述大学特性以及评估教学和学习文化影响的资料。

组织民族志依赖于深度访谈、观察和参与式观察【Participant Observation】,以便抓住源自内部的组织"感觉(feel)"、个体成员的观念和共享符号的共同体(Jones, 1996)。研究者引导参与者用他们自己的话语和表现手段来解释事件,无论是故事、庆典、仪式、谚语或是物品。他们注意个人工作空间的装饰、街头涂鸦、海报和传单艺术、照片、公告板、人们门上张贴的图画和语录、影印的传说以及其他行为和文化的物质形式。他们记录下各种口头表述,如传统的谚语、隐喻、口号、演讲、谣言、信仰、个人经历故事和广为流传的故事。他们观察(并且有时参与)风俗、仪式、人生庆典、节日活动、游戏、姿态、社会常规和大型典礼。

通过记载和分析有关组织象征的事例,民族志学者能够洞察组织中的领导风格、环境气氛、模糊和矛盾、紧张和压力、冲突以及竞争。民族志还发掘了长久以来理所当然地指导人们行动、提供决策知识以及促进或妨碍组织效能的价值和处事方式。

——Michael Owen Jones

(叶鹏飞译　高勇校)

参考文献

Dandridge, T. C., Mitroff, I., & Joyce, W. F. (1980). Organizational symbolism: A topic to expand organizational analysis. *Academy of Management Review*, *5*,77-82.

Georges, R. A., & Jones, M. O. (1995). *Folkloristics: An introduction.* Bloomington: Indiana University Press.

Jones, M. O. (1996). *Studying organizational symbolism: What, how, why?* Thousand Oaks, CA: Sage.

Tommerup, P. (1993). *Adhocratic traditions, experience narratives and personal transformation: An ethnographic study of the organizational culture and folklore of the Evergreen State College, an innovative liberal arts college.* Unpublished doctoral dissertation, University of California, Los Angeles.

Van Maanen, J. (1979). The fact and fiction in organizational ethnography. *Administrative Science Quarterly*, *24*,539-550.

正交旋转(Orthogonal Rotation)

参见旋转【Rotations】。

他者(Other, The)

"他者"这一术语应用在人文科学涉及身份、差异性、自我(selfhood)、识别和伦理等主题的论述中。"他者"既可以表示另一个人,也可以表示作为他人的自身(self)。哲学中长期存在着定位他者的兴趣,但经常会把"他者"简化为"同者(same)"或"自身"。自我发现或自我认知可以被视为一个变化的过程,自我借此认识到自身在客体上的他者性(alterity)[他性(otherness)]。例如,黑格尔指出自我意识和自由只能通过克服他者而获得。

很多论及"他者"概念的当代人文科学文献都受到伊曼努尔·列维纳斯的早期现象学著作的启发。在这里,他者超越自身而不能简化为自身。对于列维纳斯来说,他者的他性首先是一个伦理事件,并且是以无遮掩的和脆弱性的面貌呈现在我们面前。如

果我们与一个虚弱或脆弱的人有列维纳斯式的遭遇,那么这个他者就引起了我们的兴趣,我们在反思性地了解它之前已经体验到了一个伦理反应。列维纳斯说:“从一开始,与“他者”的相遇就是我对他的责任(Levinas,1998, p. 103)。”

西蒙娜·德·波伏娃(de Beauvoir, 1949/1970)在她的经典著作《第二性》(*The Second Sex*)中,通过历史、文学和神话溯源的方式,考查了女性的压迫是如何与系统对象化的父权制过程联系起来的。结果是男性被视为“正常”而女性被看作“他者”,这导致了女性社会和个人身份的丧失。在凯特·米勒特(Millett, 1970)的《性政治》(*Sexual Politics*)(1970年首次出版,后来进行过修订)一书和当代女性主义者的著作中,女性坚持表现为“他者”,来反对社会规范的男性建构。

法国后结构主义者已经发展出一系列关于他者、他性和他化(othering)这些概念的解释。他化是当人们认为自己正常而注意到别人有所不同时的作为。在米歇尔·福柯的作品中,他者是那些受害的和边缘化的人——精神病患者、囚犯、女人、男同性恋、女同性恋和有色人种。谱系分析是要复原将作为我们的他者的那些人制度化和边缘化的实践所依存的社会、文化和历史根源。

保罗·里克尔在他关于个人和叙事身份的研究中,探讨了自我与他性之间的辩证关系。他对存在于自身身体中的他性的体验特别感兴趣。里克尔提到,他性存在几种不同的层次。有“我自己作为肉体”的他性、其他人(陌生人)的他性以及道德心“召唤”上体验的他性。这些区分可能会帮助我们理解早期和当代学者的探讨过程。

在朱丽娅·克里斯蒂娃(Kristeva, 1991)的《自我中的陌生人》(*Strangers to Ourselves*)一书中,他者的概念是指我们身边的陌生人、外地人和外国人。她提出问题,如果我们不能与自身内部的他者达成妥协,又如何能与这些陌生人和平相处呢?在她的分析中,旨在呈现隐藏在我们自身中的他者的他者性。表现出藏在我们统一性表面下的他异性。雅克·德里达颇有影响力的著作是其他后结构主义者关于主体间关系研究的典型范例,他认为他者是一个主题,用以提出关于责任、义务和道德选择的问题。他说道:“如果没有牺牲其他的他者,其他的他者们,我就无法回应另一个人的召唤、请求、职责,甚至是爱。”(Derrida, 1995, p. 68)

关于他者和他性的含义问题,突出地展现了在诸如多元文化论、性别研究、酷儿理论、种族区分和后殖民研究此类领域的定性研究。在查尔斯·泰勒有关多元文化论的政治的著述中,根据有时没有共同标准的对自我认同的追求和对社会认同的复杂需求,提出了“重要他者(significant other)”的概念。在哲学中,他者的概念同样成为一个普遍关注的焦点。特别值得一提的是阿洛哈纳·林吉斯(Lingis, 1994b)关于他者的现象学研究,他翻译了列维纳斯的一些主要作品。在他的《相异者的共同体》(*The Community of Those Who Have Nothing in Common*)一书中,林吉斯展开对这一问题的探讨,即我们如何能与那些和我们没有亲属关系,没有共同的语言、宗教和经济利益的人,就文化的他者(cultural others)体验到道德责任和共同性。他在《虐待》(*Abuses*)(Lingis, 1994a)和《律令》(*The Imperative*)(Lingis, 1998)等著作中继续这条研究路线,创造出一种全新的哲学方法。在他的作品中,林吉斯糅合旅行见闻的唤醒回忆法(evocative methods of travelogue)、书信写法、伦理学,以及主题沉思法——诸如刑讯逼供、战争、身份、欲望、激情和他在印度、巴厘岛、孟加拉、拉丁美洲等地“面对面”遇到边缘化他者中的理性。

——Max van Manen

(叶鹏飞译　高勇校)

参考文献

de Beauvoir, S. (1970). *The second sex.* New York: Alfred A. Knopf. (Original work published 1949.)

Derrida, J. (1995). *The gift of death.* Chicago: The University of Chicago Press.

Foucault, M. (1995). *Discipline and punish: The birth of the prison.* New York: Vintage Books.

Kristeva, J. (1991). *Strangers to ourselves.* New York: Columbia University Press.

Levinas, E. (1998). *On thinking-of-the-other: Entre nous.* New York: Columbia University Press.

Lingis, A. (1994a). *Abuses.* Berkeley: The University of California Press.

Lingis, A. (1994b). *The community of those who have nothing in common.* Bloomington: Indiana University Press.

Lingis, A. (1998). *The imperative.* Bloomington: Indiana University Press.

Millett, K. (1970). *Sexual politics.* Garden City, NY: Doubleday.

Ricoeur, P. (1992). *Oneself as another.* Chicago: The University of Chicago Press.

Taylor, C. (1994). The politics of recognition. In C. Taylor & A. Gutmann (Eds.), *Multiculturalism and the politics of recognition* (pp. 25-73). Princeton, NJ: Princeton University Press.

离群值(Outlier)

离群值是一个远离其他值的观察值【Observation】。例如,在社区收入研究中,假设有一个百万富翁的收入为2 000 000美元,远远超过下一个最高收入150 000美元。这个百万富翁就是收入变量上的一个离群值。离群值可能导致推断问题,特别是当它们极端偏离或数量过多时。例如,在一个社区领导者的收入研究中,$N=100$,并有少数人是千万富翁,那么估计的平均值就会歪曲对典型的社区领袖收入的认识。在这种情况下,一个更好的集中趋势量度是中位数【Median】。离群值还会扭曲其他统计量,比如回归系数【Regression Coefficient】。常见的补救措施是将离群值从分析中剔除,但这种方法的缺点在于丢失了信息。离群值的转换【Transformation】可能是理想的解决方案,因为它可以纳入这些极端值,同时能够维持样本的完整。

——Michael S. Lewis-Beck

(叶鹏飞译　高勇校)

* 也可参见强影响统计量【Influential Statistics】。

P

p 值(p Value)

统计假设检验中的 p 值可以定义为概率【Probability】,是由样本数据计算出的在特定方向虚无假设【Null Hypothesis】为真的前提下,出现观察样本以及特定方向上更极端情况的概率。这个特定的方向是由单尾【One-Tailed】(或单侧【One-Sided】)的方向决定的。例如,用符号检验【Sign Test】来检验原假设“抛硬币后正反面朝上的概率相等”是否为真。我们抛 10 次,结果 9 次正面朝上。样本量就是 $n=10$,检验统计量是 S,定义为上述样本中正面朝上的次数。我们观察到的样本结果 $S=9$。这个样本结果是一个事实。问题是这个样本结果是否符合原假设中的说法,即抛硬币后正反面朝上的概率相等。如果不符合,那么这个说法一定是错误的,因此可以拒绝原假设。为了确定样本结果是否与虚无假设一致,需要计算当抛硬币后正反面朝上的概率相等的情况下,抛 10 次硬币中有 9 次或更多次正面朝上的概率。这个概率由 $n=10$ 和 $p=0.50$(当虚无假设为真时,硬币正面朝上的概率为0.50)的二项分布计算如下

$$
\begin{aligned}
p\text{ 值} &= p(S \geqslant 9) \\
&= \sum_{t=9}^{10} \binom{10}{t} (0.50)^{t}(0.50)^{10-t} \\
&= 0.002\,0
\end{aligned}
$$

由计算结果得知虚无假设为真的概率是 0.002 0,这是一个很小的概率。因此,需要拒绝虚无假设,认为抛硬币后正反面朝上的概率是不等的,硬币似乎正面朝上的可能性更大,即 $p>0.50$。

传统经典的假设检验方法是预先设定一个数值 α,即犯第Ⅰ类错误【Type Ⅰ error】的概率,根据 α 值判定 S 是否属于拒绝域;例如,如果 $S \geqslant c$ 则拒绝虚无假设(c 为常数)。这个预先设定的 α 值通常是任意的(习惯上选择 0.05),但它对最后的结果判断有很大影响。

报告 p 值是虚无假设检验结果的一种表达方式。与只报告在预先选定的 α 水平上是否拒绝虚无假设的确定性结论相比,研究者报告 p 值可以让读者或用户基于此 p 值的大小作出自己的判断。读者可以选择他们自己的 α 水平,若 $p \leqslant \alpha$,拒绝虚无假设;反之,不拒绝。

如果备择假设【Alternative Hypothesis】不是一个特定方向的预测,即选择的是双尾【Two-Tailed】(或双侧【Two-Sided】),那么 p 值的定义就不是唯一的。实际上,p 在两个可能的方向上有不同的值。其中较小的 p

值是有意义的。因此,我们既可以报告比较小的那个 p 值,表示是单侧的 p 值,也可以将这个单侧 p 值乘以 2。

—Jean D. Gibbons

(叶鹏飞译 高勇校)

配对相关(Pairwise Correlation)

参见相关【Correlation】。

成对删除(Pairwise Deletion)

成对删除这一术语应用于计算机软件程序(如 SPSS)对于缺失数据【Missing Data】的处理过程中。缺失值的成对删除是指仅删除与分析相关的有缺失值的样本。考虑以下的情景:我们有 200 个样本和 10 个变量,现在想要进行皮尔逊相关分析【Pearson's Correlation Coefficient】。我们先取前三个变量。变量 1 有 4 个缺失值,变量 2 有 8 个缺失值,变量 3 有 2 个缺失值。假设每个样本缺失的变量都不同,即变量 1 的 4 个缺失值和变量 2 的 8 个缺失值没有重合。(这种情况在实践中不常见,因为有些参与者可能同时缺失变量 1 和变量 2 的值,即变量 1 和变量 2 的缺失值可能出现在同一个样本中)当我们分析 10 个变量时,变量 1 和变量 2 之间的相关分析是基于 188 个样本的分析(因为在变量 1 和变量 2 中有 12 个缺失值)。变量 1 和变量 3 的相关分析是基于 194 个样本的分析(因为这两个变量中有 6 个缺失值)。以此类推,变量 2 和变量 3 的相关分析是基于 190 个样本的分析(因为这两个变量中有 10 个缺失值)。

——Alan Bryman

(叶鹏飞译 高勇校)

* 也可参见删除【Deletion】、成列删除【Listwise Deletion】。

定群(Panel)

定群指的是对同一分析单位进行至少两次数据收集的研究设计【Research Design】。通常情况下,分析单位为调查【Survey】中的某一个体。在两层定群研究中,会对 t 时刻的响应对象在 $t+1$ 时刻再次访问;同样,在三层定群研究中,将会在 $t+2$ 时对其再一次访问。有时,定群中的分析单位具有一定聚集性,例如国家。比如,来自欧洲某些国家的样本可以在 t 时进行测量后又在 $t+1$ 时再一次测量。定群研究一个特殊价值在于它考虑到因果【Causal】推断中的时间性,因为潜在的因果变量 X 能够在 Y 发生前真正地测量到。除此之外,定群研究可以校正社会因素的改变。例如,通过一段长时间的重复访谈【Interviewing】,定群研究(不同于同期群【Cohort】设计)可以区分年龄效应、周期效应【Period Effects】以及同期群效应。

定群研究的一个主要缺点是在不同时点的研究中存在数据损耗【Attrition】,特别是在定群研究中个体的退出。由此带来的

问题是人口特征随时间而变化,或许会导致原始的定群已经无法代表最初抽样【Sample】的总体【Population】,这也是定群研究相关的一个问题。例如,随着移民的迁入,定群需要新的样本加入才能反映改变后的人群。

——Michael S. Lewis-Beck

(叶鹏飞译　高勇校)

定群数据分析(Panel Data Analysis)

定群数据是指对每个抽样单元进行多次观察得到的数据集。这样的数据可以通过对一系列横断面单位进行的时间序列观察得到,横断面单位可以为国家、州、地区、公司或随机抽取到的个人或家庭。研究包含了纵向数据分析,其重点在于研究个体的历史。两个著名的美国定群数据分别是收入动态定群研究(Panel Study of Income Dynamics, PSID),由密歇根大学社会研究学院收集;全国劳动力市场经验纵向调查(National Longitudinal Surveys of Labor Market Experience, NLS),由俄亥俄州立大学人力资源研究中心创立。通过相关网址可以访问使用定群数据的国家级研究项目列表,包括比利时家庭定群调查、荷兰社会经济定群研究、卢森堡家庭定群研究以及欧洲社区家庭定群研究。收入动态定群研究创始于1968年的4 802个家庭,并包含了过度抽样的贫困家庭。每年进行的访问记录了每个家庭的社会经济特征以及在这些家庭中的大约31 000个体,其中收集的全部变量超过5 000个。NLS跟踪了五种不同劳动力人群。最初的样本包括了1996年45~59岁男性5 020名,1966年14~24岁男性5 225名,1967年30~44岁女性5 083名,1968年14~24岁女性5 159名,以及1979年14~21岁青年12 686名。在青年研究中存在对黑人、西班牙裔、贫穷白人以及军人的过度抽样。所收集的变量高达数千。定群数据集的例子还有美国当前人口调查(Current Population Survey, CPS),这是由美国人口统计局开展的全国家庭月度调查。CPS统计了失业率和其他劳动力统计数据。与NLS跟PSID这两个数据集相比,CPS包含的变量较少,时间跨度更短,并且没有纳入流动人口。然而,它覆盖了更多的样本,能够代表所有人口统计群体。

1986年,萧政提出了关于定群数据的优势以及局限性。明显的优点包括了一个较大的数据,因为定群数据包含了对同一个体多次观察的结果,这意味着变量当中相比传统的横断面研究与时间序列研究存在更多的可变性以及更少的共线性【Collinearity】。例如,在一个已知货物(如汽油)的需求方程中,货物的价格以及收入可能在某一国家的年度观察期内会有非常强的相关性。然而,通过将不同国家的数据堆积或合并,数据的变异得到增加,而共线性将会减弱。另外,利用更丰富的数据,我们可以得到更可靠的估计和测试更复杂的行为模型,模型假设的限制也可以放宽。定群数据的另一个优势在于控制个体的异质性的能力。没有控制那些未被观察到的个体效应会导致估计结果时偏差的产生。例如,在一个收入方程中,一个人的工资是基于各种个体属性的回归估计,如教育、经验、性别、种族等,但是误差项可能还包括未被观察到的个人特征,如能力,并与一些解释变量相关,比如教育。横断面研究通过收集难以获得的双胞胎资料,企图对这种未被观

察到的信息加以控制,然而通过定群个体资料,我们可以对个体进行多次测量,然后消灭个体中未被观察到的不变的特征。定群资料的另一个优势是可以更好地识别并且估计那些在单纯的横断面研究与单纯的时间序列数据中难以探测到的效应。尤其是定群数据能够更好地研究较为复杂的动力学行为的问题,例如,通过横断面研究,可以估计某一个时点的失业率;而重复的横断面研究可以得到失业率在一定时期内的变化;只有定群研究可以估计在某一时点失业的人群在另一时点仍然失业的比例。

定群数据集的局限性主要有几个方面:设计中的问题、数据收集问题以及定群研究中数据管理的问题(参见 Kasprzyk, Duncan, Kalton, & Singh, 1989)。这些问题包括:涵盖范围问题(所关心的人群没有完全纳入)、无响应问题(由于缺乏合作或采访失误造成的无应答)、回忆问题(被访问者回忆出现偏差)、采访频率问题、采访时间跨度问题、参考期问题、防止回忆期外事件被纳入回忆期内的问题,以及抽样时机引起的偏倚。定群数据的另外一个局限性是来自测量误差,不清晰的问题、回忆错误、故意给出错误信息、不合适的被调查者、信息的错误记录以及采访者效应都可能造成测量误差的放大。尽管这些问题也可能发生在横截面的研究中,但是在定群数据的研究中会更突出。定群数据集还可能会由于样本选择问题而呈现出偏倚。在初始的定群中,被调查可能会拒绝参与,或者访问者在家中找不到人,这样的样本可能会导致在推断中存在偏倚。尽管这种无应答也能发生在横断面研究中,但是在定群研究中却更为严重,因为后续的定群调查仍然会被这样的样本所限制。受访者可能会去世、搬家或者发现应答的代价过大,最终可能失访。不同定群的损耗率是不一样的,通常情况下会是后一次采访比本次高,但增长速率会下降。经典的定群会每年都去访问每个人,询问的却只是一年中某一小段时间中的情况。这意味着只有访问的人数接近无穷大时,相应的估计才能渐进准确。增加定群的时间跨度是无法不计成本的。实际上,每一次新的定群都增加了损耗的机会,同时也增加了估计定群数据模型中定性有限因变量的计算难度。(Baltagi, 2001)

尽管随机系数模型【Random-Coefficient Model】可以应用于估计以及设定定群数据模型(Hsiao, 1986),大多数定群数据模型采用的是有多个误差扰动项的简单回归形式,如下所示:

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \mu_i + v_{it}, i = 1, \cdots, N; t = 1, \cdots, T$$

方程中 y_{it} 代表了第 i 个个体在时间 t 的工资对数值,$\mathbf{x}_{it}$ 是一个表示观察了 k 个解释变量的向量,包括教育、经历、种族、性别、婚姻状态、工会成员和工作时间等。此外,$\boldsymbol{\beta}$ 是一个表示未知系数的 k 向量,μ_i 为未观测到的个体效应,v_{it} 为一个均值为 0,方差为 σ_v^2 的随机干扰项。误差因素遵循单因素方差分析(ANOVA)。如果 μ_i 表示需要估计的固定参数,这个模型就被称为固定效应模型【Fixed-Effects(FE)Model】。$\mathbf{x}_{it}$ 被假定为对于所有的 i 和 t 都是独立于 v_{it}。这里的推断取决于特定的 N 个观测个体。这种情况实际上相当于引入($N-1$)个虚拟变量(individual dummies)来估计这些个体效应。这导致了自由度的大量损失,并加重了回归量之间的多重共线性【Multicollinearity】问题。而且,对于 N 很大的定群数据,这在计算上是不可行的。在这种情况下,我们可以通过对 $\bar{y}_{it} = y_{it} - \bar{y}_i$ 在同样定义的 $\bar{\mathbf{x}}_{it}$ 上运用最小二乘,来消掉 μ_i 估计 $\boldsymbol{\beta}$ 值,这里字母上的点表示该指标上的总和,横条表示求均值。

这种转换被称为内转换，相应的估计量 $\boldsymbol{\beta}$ 被称为内估计值或者固定效应估计值。必须注意，固定效应估计量无法估计任何不随时间变化的变量效应，如性别、种族、信仰或工会参与。这些变量在内转换中没有被纳入。当需要考虑这些变量对于收入的影响时，这就是一个主要的弊端。忽略个体未被观察到的效应（例如不引入虚拟变量进行最小二乘）会导致偏倚的出现，得到的回归系数也并非一致估计量。

如果 μ_i 代表了均值为 0，方差为 σ_μ^2 的独立随机变量，这个模型就称为随机效应模型。模型前面的部分取决于 $\mathbf{x}_{it}$。另外，μ_i 与 v_{it} 被假定为条件性独立。随机效应（RE）模型可通过广义最小二乘法（GLS）进行估计，可通过对 $y_{it}^* = y_{it} - \theta \bar{y}_i$ 在同样定义的 $\mathbf{x}_{it}^*$ 上运用最小二乘回归来获得，这里的 θ 是方差成分 σ_μ^2 和 σ_v^2 的一个简单函数（Baltagi，2001）。相应的 GLS 估计量 $\boldsymbol{\beta}$ 就是 RE 估计量。对于这个 RE 模型，我们可以估计个体恒量变量（individual-invariant variables）的效应。方差成分的最佳二次无偏（BQU）估计量是基于真实干扰的方差分析型估计量（ANOVA-type estimators），并且在干扰是正态的情况下，它们也是最小方差无偏（MVU）估计量。通过使用 OLS 或者固定效应残差替代真实干扰，我们可以获得方差成分的合适估计值。对于随机效应模型，OLS 仍然是无偏且一致的，但并不有效。

固定效应与随机效应何者更优，在生物统计学和计量经济学文献中对此有过激烈争论。在一些应用中，随机效应与固定效应得到了不同的估计结果，尤其是当 T 很小而 N 很大时。豪斯曼（Hausman，1978）根据两者估计的差异开展了一项设定检验，虚无假设是个体效应与 $\mathbf{x}_{it}$ 是不相关的。这个检验的基本思想是固定效应的估计量 $\bar{\boldsymbol{\beta}}_{\mathrm{FE}}$ 是一致的，无论个体的效应与 $\mathbf{x}_{it}$ 是否相关。这是正确的，因为由 $\tilde{y}_{it}$ 描述的固定效应转换将 μ_i 的效应排除在模型之外。实际上，这是对固定效应模型的现代计量经济学的解释，μ_i 是随机的，但是却与所有的 $\mathbf{x}_{it}$ 相关。然而，如果虚无假设为真，则固定效应估计量在随机效应的设定下并不有效，因为它仅仅依赖于数据中个体的内部变化。另一方面，随机效应估计量 $\bar{\boldsymbol{\beta}}_{\mathrm{RE}}$ 在虚无假设前提下是有效的，但是当效应与 $\mathbf{x}_{it}$ 相关时，它存在偏倚以及不一致性。两个估计量的差值 $\hat{q} = \bar{\boldsymbol{\beta}}_{\mathrm{FE}} - \hat{\boldsymbol{\beta}}_{\mathrm{RE}}$ 在虚无假设成立时趋向于 0，而在备择假设中不为 0。在虚无假设成立时该差值的方差等于方差之差，即 $\mathrm{var}(\hat{q}) = \mathrm{var}(\bar{\boldsymbol{\beta}}_{\mathrm{FE}}) - \mathrm{var}(\bar{\boldsymbol{\beta}}_{\mathrm{RE}})$，因为 $\mathrm{cov}(\hat{q}, \hat{\beta}_{\mathrm{RE}}) = 0$。Hausman 检验统计量是基于 $m = \hat{q}'[\mathrm{var}(\hat{q})]^{-1}\hat{q}$，并且在虚无假设前提下，渐近地服从自由度为 k 卡方分布。

对于定群模型估计的最大似然和广义矩方法，请参考巴尔塔吉的著作（Baltagi，2001）。由于篇幅所限，对于定群数据模型的讨论并没有展开，包括缺失值的处理、动态分析、测量误差、定性受限因变量、内生变量以及回归变量非平稳性。但是，我们还是要讨论一些经常遇到的特定的定群数据集，例如伪定群与轮换定群。伪定群指的是定群数据来源于横断面的重复，尤其是在某些国家中不存在定群但是存在多次独立的研究。例如，英国家庭支出研究调查了每年大约 7 000 家庭。这些都是独立的调查，因为无法做到像定群研究一样，追踪到在调查过程中同样的家庭。但是，我们可以追踪同期群并基于同期群均值来估计经济关系。伪定群中不存在使真正的定群研究头痛的样本损耗问题，并且可以历时更长。一个重要的问题是关于同期群的最佳数量问题。同期群越多，每个同期群的规模就越小，从中抽取出来的样本

也就越少。相对的,同期群越少,样本同期群的均值估计就越准确,但是定群的有效样本量却会因此越少。

轮换定群试图通过取代在每一时期都会存在一部分的缺失家庭的方法来保持同样数量的家庭。这种方法在一个高损耗的定群研究中是非常必要的。轮换定群可以让研究者对抽样时机引起的偏倚效应的存在进行检测。如果我们期望得到相同响应时,两次不同时点进行的访谈结果却有着显著差异,这就表明存在抽样时机引发的偏倚效应。

随着越来越多地使用多个时点的跨国数据来研究购买力平价(purchasing power parity)、增长收敛性和国际研发溢出效应,定群数据计量经济学的焦点转移到研究拥有大量 N(国家数量)和大 t(时间序列的长度)的宏观定群数据的渐近性质,而不是通常的那种拥有大量 N 和小 t 的微观定群数据的渐近性质。研究人员认为,变量的时间序列成分如人均国内生产总值的增长具有强烈的非平稳性。非平稳定群得到的一些特殊结果是,很多检验统计量和估计量存在正态极限分布。这与文献中非平稳时间序列研究相反,那些文献中的极限分布是韦氏过程(Weiner processes)的复杂函数。几个应用于时间序列文献的单位根检验已经扩展到了定群数据(Baltagi,2001)。然而,也有人对使用定群数据的方法进行了批评,认为定群数据单位根检验不能解决购买力平价(PPP)。实际上,定群数据中 PPP 的结果是依据研究的国家、研究时期以及单位根检验类型三者进行混合的结果。更为严重的问题是,对于购买力平价,定群数据检验并不能解决单一时间序列中单位根检验低效的问题。毕竟,单一单位根的虚无假设有别于定群单位根关于购买力问题的虚无假设。同样的,定群单位根检验并没有解决各国经济增长收敛性问题。然而,它有助于促进对动态定群数据模型的深入研究。

在过去的 20 年里,定群数据方法文献表现出显著的增长。我们还不能涉及众多理论和经验上的最新贡献。篇幅所限,不能讨论很多有价值的贡献。有些主题仍然在起步阶段,但是发展迅速,比如非平稳定群的使用,半参数和非参数方法在定群数据中的应用。希望这一介绍会激起读者的兴趣并鼓励更多关于此类主题的阅读。

——Badi H. Baltagi

(叶鹏飞译　高勇校)

参考文献

Baltagi, B. H. (2001). *Econometric analysis of panel data.* Chichester, UK: Wiley.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, *46*, 1251-1271.

Hsiao, C. (1986). *Analysis of panel data.* Cambridge, UK: Cambridge University Press.

Kasprzyk, D., Duncan, G. J., Kalton, G., & Singh, M. P. (1989). *Panel surveys.* New York: John Wiley.

范式(Paradigm)

在日常使用中,范式是指要遵从的一个模型或一个范例,或是指一种做事的现成体系或方式。这一概念是由托马斯·库恩(Kuhn,1970)在他讨论科学进步的本质时引入科学哲学。

库恩反对合宜的科学方法可以事先确定这样的科学哲学观点,如波普尔的证伪主义【Falsificationism】。库恩(Kuhn,1970)聚焦

于科学家共同体的实践。他发现这一共同体共享一种范式或者"学科基体(discipline matrix)",这包括他们研究的现实本质是什么(他们的本体论【Ontology】),包括组成的要素以及它们是如何相互关联的;适合调查这一现实的方法(他们的认识论【Epistemology】)以及过去科学成就的公认范例,这些范例既提供了深入实践的基础,也为未来希望成为共同体成员的学生提供了模型。他认为一门成熟科学是由单一范式所主导的。

根据库恩(Kuhn,1970)的观点,科学家在大多数时间里从事常规科学(normal science),研究被"解决谜题(puzzle-solving)"活动和严格基于范式的假设和规则所左右。常规科学通过检验它的预测从而进一步阐明和丰富其推论扩展了范式中的知识;它不以发现意想不到的新奇事实或理论为目标。在常规科学的过程中,范式不会受到挑战或者检验;未能解决一个难题将被视为科学家的失败,而不是范式的失败。

时常会出现一些无法解决难题的时候,或范式的预期与现实观察的差异。最初这些异常可能被忽略,因为恪守范式会对他们的认识产生内在阻力。库恩(Kuhn,1970)认为范式是感知的先决条件,我们的理解不仅依赖于我们看见的,还有赖于我们以前视觉经验和概念经验,它们会教我们去看哪些东西。坚持范式类似于坚持一种信仰,认为其中有问题的话就会被视作异端。

异常现象可能会导致范式的信任危机。伴随着竞争性学说的扩散、尝试任何事的意愿、不满的表达、对哲学的求助和对基本原理的争论,出现了一个时期的非常科学(extraordinary science)。这种情况为新范式和新理论的出现作好准备。

库恩(Kuhn,1970)描述了一个新范式取代旧范式的科学革命(scientific revolution)过程。新范式可能被提议取代现有的范式——新的范式可解决异常现象引起的新难题,也可处理以前范式已解决的难题。然而,这样的革命发生很缓慢,需要一代人或许更久的努力。根据库恩的说法,科学家从旧的范式转变到新的范式的过程类似于宗教的皈依;它不仅需要采用一种完全不同的观察世界的方式,而且还要生活在一个不同的世界。一旦新的范式建立起来,常规科学的新阶段将会出现。最后,新的异常现象将会出现,而进一步的革命也将发生。

库恩(Kuhn,1970)认为竞争的范式是不可通约的。这是因为一个科学共同体提出的观念和理论主张是基于对他们有特殊意义的范式中的假设和信念而产生的。由于范式体现不同的和不兼容的世界观,它很难在不同的科学共同体进行有效沟通,进而就不可能就相互竞争的理论进行裁决。因为它们之间没有中立的观察语言,没有共同的词汇,没有解决争端的公共平台。

对于库恩来说,科学进步不是通过观察(归纳【Induction】)或者通过对新假设的决定性检验(演绎/证伪主义)来概括积累的——它是通过科学革命改变科学家共同体看待世界以及定义和解决难题的方式来实现的(参见 Blaikie, 1993, pp. 105-110; Riggs, 1992, pp. 22-59)。

库恩的成果催生了大量的研究文献,并且受到来自科学哲学家和科学史学家的详细批评,例如卡拉托什和劳丹(综述参见 Riggs, 1992)。在 1960 年代和 1970 年代,它被作为理解社会学危机的一个框架(参见,如 Friedricks, 1970),那时在各个竞争性范式之间存在着激烈的争论。

——Norman Blaikie

(叶鹏飞译　高勇校)

参考文献

Blaikie, N. (1993). *Approaches to social enquiry*. Cambridge, UK: Polity.

Friedricks, R. (1970). *A sociology of sociology*. New York: Free Press.

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.

Riggs, P. J. (1992). *Whys and ways of science*. Melbourne, Australia: Melbourne University Press.

Ritzer, G. (1980). *Sociology: A multiple paradigm science*. Boston: Allyn & Bacon.

参数(Parameter)

参数是一个揭示总体【Population】某些特征的量。总体是包含研究者感兴趣的所有结果的集合。然而,在统计学中,总体被更技术性地定义为从一次随机试验【Experiment】中得到的所有可能结果的集合。例如,随机试验可能由对美国公民个人的纳税观察构成。总体包含每个美国公民的税单。总体就可以被一个随机变量【Random Variable】X(例如纳税)及其相关概率分布(例如纳税的概率函数或密度函数)所描述。当我们清楚随机变量 X 的概率【Probability】分布 $f(x)$ 时,总体被认为是已知的。如果总体已知,参数可以很容易从总体数据中计算出来。因此,参数是代表总体特征的固定量。

描述性分析的两个常用参数是总体的均值【Mean】和方差【Variance】。总体均值显示总体的数值平均,而总体方差表示观测值偏离均值的平均平方偏差。例如,在上述税收试验中,有美国公民的平均税收和税收方差。其他参数的例子可以是总体的偏度【Skewness】、峰度【Kurtosis】、相关性【Correlation】和回归偏斜率参数(regression partial slope parameter)。

参数没有变异性,因为它们可以从一个已知总体计算出来。然而,通常使用总体是不方便或不可能的。总体可能会相当大,如上述纳税人的例子,获得或输入这些数据太过困难。在社会科学中获取总体往往很困难。对研究者来讲,这个数据可能并不能直接获取,或根本不存在。

在社会科学研究的大部分时候,总体是不可获得的,所以参数是一种未知的理论构念。参数存在于真实世界的某个地方,但是研究者必须从一个样本中估计其值。在样本数据估计参数的过程中,样本中所有条目可以看成是通过随机试验从总体中获取的对随机变量(random variables)的观测。它们被认为有相同的概率分布并相互独立地被抽取。样本统计量(sample statistics)揭示样本的某些特征,它们通过这些随机变量来估计总体参数。例如,在上述税收试验中,样本均值可以通过将所有样本条目的样本观察值求和,再与它们各自的概率相乘而获得。这被认为是总体均值的一个估计量。

——B. Dan Wood
Sung Ho Park
(叶鹏飞译　高勇校)

参考文献

Blalock, H. M., Jr. (1979). *Social statistics*. New York: McGraw-Hill.

Casella, G., & Berger, R. L. (2001). *Statistical inference* (2nd ed.). New York: Wadsworth.

Spiegel, M. R., Schiller, J., & Srinivasan, R. A.

(2000). *Schaum's outline of probability and statistics* (2nd ed.). New York: McGraw-Hill.

参数估计(Parameter Estimation)

参数估计是要从样本统计量中发现总体参数的值。总体参数是固定的并且一般是未知量。如果我们知道某一总体的所有可能条目及它们各自的概率【Probability】,我们能够计算总体的参数值。然而,我们通常只有总体中的某一样本,所以我们必须通过样本统计量估计总体参数。体现样本特征的样本统计量(sample statistics)可被认为是相应总体参数的估计量【Estimator】。注意,因为样本统计量是基于对总体的抽样样本,所以样本统计量是随机变量【Random Variable】。

样本统计量用于对固定的总体参数的估计。例如,样本均值可作为总体均值的估计值。然而,由于抽样误差【Sampling Error】的存在,并不能保证依据统计量得出的所有参数估计都能准确地反映总体参数。然而,我们能通过好的估计量最小化参数估计的不确定性。一个好的估计量应满足一定的统计特性,例如无偏性【Unbiasedness】、有效性【Efficiency】、渐近有效性或者一致性。前两个属于有限样本的性质,后两个属于渐近性质。

无偏性要求估计量的期望值应该等于总体参数的真值。即无偏性要求 $E(\hat{\theta})=\theta$,θ 是总体参数,$\hat{\theta}$ 是参数估计值。再次注意,这并不意味着一个特定的估计量将等于参数的真值。相反,它体现重复抽样的特征。我们期望来自重复抽样样本的估计值均值等于参数的真值。有效性要求估计量的方差小于或等于其他相同参数的无偏估计量。即 $\mathrm{Var}(\hat{\theta}_1) \leqslant \mathrm{Var}(\hat{\theta}_2)$,$\hat{\theta}_2$ 是来自除 $\hat{\theta}_1$ 外的任一估计量。有时有偏的估计量却比无偏的估计量的方差小。在这种情况下,可采用最小均方误差【Mean Squared Error, MSE】标准。计算公式如下:$\mathrm{MSE}(\hat{\theta}\mid\theta)=\mathrm{Var}(\hat{\theta})+\mathrm{Bias}(\hat{\theta}\mid\theta)^2$。不考虑偏倚,具有最小均方误差的估计量通常被认为是更佳的估计量。

一个好的估计量能满足无偏性和有效性的要求。然而,小样本的估计量难以达到这些要求,但随着样本量的增加,估计量可以获得理想的统计特性。这样的估计量就具有渐近有效性或一致性。渐近有效性(asymptotic efficiency)意味着估计量的渐近分布的方差小于其他渐近无偏估计量的方差。一致性(consistency)意味着当样本容量趋于无穷大时,估计量依概率收敛于总体参数的真实值。这两个性质相当于有限样本中的有效性和无偏性。

——B. Dan Wood
Sung Ho Park
(叶鹏飞译　高勇校)

参考文献

Blalock, H. M., Jr. (1979). *Social statistics*. New York: McGraw-Hill.

Casella, G., & Berger, R. L. (2001). *Statistical inference* (2nd ed.). New York: Wadsworth.

Judge, G. G., Hill, R. C., Griffiths, W. E., Lutkepohl, H., & Lee, T. (1988). *Introduction to the theory and practice of econometrics*. New York: John Wiley.

Spiegel, M. R., Schiller, J., & Srinivasan, R. A. (2000). *Schaum's outline of probability and statistics* (2nd ed.). New York: McGraw-Hill.

部分相关(Part Correlation)

部分相关系数 $r_{y(1\cdot2)}$,也被海斯(Hays,1972)称为半偏相关系数,与偏相关系数 $r_{y1\cdot2}$的不同在于:它能将控制变量 X_2的影响从变量 X_1中去除而不从 Y 中去除。

偏相关【Partial Correlation】分析中的三个步骤现在被缩减为两步:第一步,计算 X_1 在 X_2上的二元回归【Regression】分析的残差 $X_1-\hat{X}_1$;第二步和最后一步,计算残差 $X_1-\hat{X}_1$ 与原始 Y 之间的零阶【Zero-Order】相关系数,即部分相关系数 $r_{y(1\cdot2)}$。对于高阶【Higher-Order】系数,如 $r_{y(1\cdot23)}$,以同样的方式,将控制变量 X_2和 X_3的影响从变量 X_1中去除而不是从因变量 Y 中去除。一个关于 $r_{y(1\cdot2)}$的例子是,控制学生的学习成绩(X_2)后,课外活动(变量 X_1)和学生声望(变量 Y)之间的相关性。为阐明部分相关的原理,我们以 5 名学生的简单样本(虚构)为例,以上三个变量值如下,Y :1,3,2,6,4;X_1:0,1,3,6,8;X_2:4,4,1,2,0。通过残差 $X_1-\hat{X}_1$ 的值-0.70,0.30,-2.53,2.08,0.86 和原始的 Y 值 1,3,2,6,4 的零阶相关系数计算得到部分相关系数 $r_{y(1\cdot2)}=0.83$。这就是声望与课外活动之间的相关性,其中已经排除了学习成绩对课外活动的影响。

在处理相关内容时,部分相关应用较少。当从因变量 Y 中去除控制变量 X_2的影响时,因变量 Y 的方差基本消失,以至于几乎没有进一步去解释的必要,而这是部分相关在实证研究中优于偏相关的原因之一。当然,控制变量在这个例子中很特别,因为这个控制变量几乎完全与其中一个考虑的变量共享了它的方差。

一个关于部分相关系数更合理运用的例子是科尔森(Colson,1977, p. 216)的研究,关注的是一个国家样本中的平均年龄(A)对粗出生率(B)的影响。但是粗出生率不独立于年龄结构,反之亦然,因为平均年龄 A 也包括生育成分,由此出现了混杂问题。如果我们纳入总和(周期)生育率作为一个变量,它可以通过总的年龄别生育率(F)来衡量并独立于年龄结构,那么这个问题可以有两种方式来解决:第一种方式,计算 A 和 F 之间的相关系数(即平均年龄和生育率的相关);第二种方式,也就是作者所选的方式,计算粗出生率(B)和排除生育影响的年龄(移除 F 影响后 A 的残差值)之间的部分相关系数 $r_{B(A\cdot F)}$。

这里去除混杂变量后不清楚哪部分变量被保留,这样的案例适合采用部分相关分析。如果可以,应该在原始概念(年龄结构和出生率)之间保持概念性距离以避免混杂。

偏相关和部分相关的计算

在上文中,我们试图依据内容来解释部分相关。然而它的作用显然是本身固有的。许多公式可以用来得到部分相关系数,并且对于每一个公式,推算出的偏相关系数都是等价的。这里我们将做一些 $r_{y(1\cdot23)}$的推导,其中 Y 与控制 X_2和 X_3后的 X_1相关。

部分相关的平方即为将 X_1加入$\{X_2,X_3\}$时对 Y 的总的解释方差比例的增量:

$$r^2_{y(1\cdot23)}=R^2_{y\cdot123}-R^2_{y\cdot23}$$

按如下公式,更便于理解:

$$R^2_{y\cdot123}=R^2_{y\cdot23}+r^2_{y(1\cdot23)}$$

多重决定系数 $R^2_{y\cdot123}$是指因变量 Y 被所

有自变量 X_1，X_2 和 X_3 解释的方差比例。它由两部分组成：第一部分 $R^2_{y\cdot23}$，是因变量 Y 由 X_2 和 X_3 解释的方差比例。第二部分是指排除了 X_2 和 X_3 的影响后仅仅 X_1 解释的方差比例，即为部分相关系数的平方 $r^2_{y(1\cdot23)}$。

如果 X_1 与 X_2 和 X_3 都不相关，那么 R^2 中的增量等于 r^2_{y1}。另一方面，如果 X_1 与 X_2 和 X_3 的相关性很强，那么部分相关的平方就与原平方相关 r^2_{y1} 完全不同。

如果变量 X_2 和 X_3 解释了 Y 方差的很大比例，X_1 能够解释的便所剩无几。在这种情况下，部分相关很小并且在控制变量大部分不能解释的情况下几乎没有可比性。因此，将 R^2 中的"绝对"增量除以不能被 X_2 和 X_3 解释的那部分 Y 的方差，我们得到偏相关系数的平方，而这等于将 X_1 加入 $\{X_2, X_3\}$ 中后，Y 的可解释的那部分方差比例的"相对"增加量：

$$r^2_{y1\cdot23}=\frac{R^2_{y\cdot123}-R^2_{y\cdot23}}{1-R^2_{y\cdot23}}$$

因此，偏相关等于部分相关的平方除以控制变量不能解释的部分：

$$r^2_{y1\cdot23}=\frac{r^2_{y(1\cdot23)}}{1-R^2_{y\cdot23}}$$

于是

$$r_{y1\cdot23}=\frac{r_{y(1\cdot23)}}{(1-R^2_{y\cdot23})^{\frac{1}{2}}}$$

显著性检验

部分相关系数和偏相关系数的显著性检验与回归分析中偏回归系数的显著性检验是一致的。

——Jacques Tacq

（叶鹏飞译　高勇校）

参考文献

Blalock, H. M. (1972). *Social statistics*. Tokyo: McGraw-Hill Kogakusha.

Brownlee, K. (1965). *Statistical theory and methodology in science and engineering*. New York: John Wiley.

Colson, F. (1977). *Sociale indicatoren van enkele aspecten van bevolkingsgroei*. Doctoral dissertation, Katholieke Universiteit Leuven, Department of Sociology, Leuven.

Hays, W. L. (1972). *Statistics for the social sciences*. New York: Holt.

Tacq, J. J. A. (1997). *Multivariate analysis techniques in social science research: From problem to analysis*. London: Sage.

偏相关(Partial Correlation)

偏相关是两个变量在其他一个或多个变量的效应保持不变的条件下的一种相关。一个经常被引用的例子是：一个地区样本中鹳的数量与婴儿数量的相关性，当乡村化程度保持不变时，这种相关就消失了。这意味着在鹳的数量上有差别的地区，同样也存在出生率的差别，但这种共变完全是由于乡村化的不同。对于乡村化程度评分相等的地区，鹳和婴儿之间的这种关系应该就会消失。这种相关就被称为"虚假"相关。

相关系数不一定会因此变为零，而是可能发生方向上的变化，我们举一个偏相关的

例子,即在控制学生的学习成绩(变量 X_2)之后,学生课外活动(变量 X_1)与学生声望(变量 Y)之间的相关性。为了说明偏相关的原理,我们给出一个虚拟的样本,五个学生三个变量的得分如下。

表 1

Y	X_1	X_2
1	0	4
3	1	4
2	3	1
6	6	2
4	8	0

为了计算偏相关系数,我们排除 X_2 对 X_1 和 Y 的影响,这样偏相关系数就变成了关于残差项 $X_1-\hat{X}_1$ 与 $Y-\hat{Y}$ 的简单零阶【Zero-Order】相关。这个原理主要有以下三步:

第一,进行 X_1 在 X_2 上的二元回归分析。计算出估计值 $\hat{X}_1$,且由 $X_1-\hat{X}_1$ 之差得到残差值,不受控制变量 X_2 的干扰。

第二,进行 Y 在 X_2 上的二元回归分析。计算出估计值 $\hat{Y}$,并得到残差 $Y-\hat{Y}$。

第三,$X_1-\hat{X}_1$ 与 $Y-\hat{Y}$ 是相关的。这两个残差间的零阶相关就是偏相关系数 $r_{y1 \cdot 2}$(即变量 Y 和 X_1 在去除变量 X_2 干扰后的零阶积矩相关系数)。

这个偏相关系数因为只控制了一个变量,所以是一个一阶系数。这个模型可以针对多个控制变量扩展到高阶【Higher-Order】偏相关。在三阶系数 $r_{y1 \cdot 234}$ 中,Y 与 X_1 的偏相关系数是在控制变量 X_2,X_3 和 X_4 之后计算的,第一步不是一个二元回归分析而是一个 X_1 对 X_2,X_3 和 X_4 的多元回归分析,得到残差项 $X_1-\hat{X}_1$。第二步是 Y 对 X_2,X_3 和 X_4 的多元回归分析,得到残差项 $Y-\hat{Y}$。然后 $X_1-\hat{X}_1$ 与 $Y-\hat{Y}$ 的零阶相关就是相关系数 $r_{y1 \cdot 234}$。

对于总体,我们用希腊字母代替拉丁字母。样本的相关系数是 $r_{y1 \cdot 2}$ 和 $r_{y1 \cdot 234}$,这样总体相关系数就可以写成 $p_{y1 \cdot 2}$ 和 $p_{y1 \cdot 234}$。

偏相关的计算

所有上述操作都汇集于表 2 中。

在这里我们可以看出,偏相关分析的原理不应该单纯地理解为当控制第三个因素时一种关系的消失。在这个案例中,Y 与 X_1 的原始相关是 $r_{y1}=0.75$,而偏相关是 $r_{y1 \cdot 2}=0.89$。所以,课外活动与学生声望之间的关系初看起来就很强烈,而当控制了学习成绩之后则变得更强。

表 2

第一步:$\hat{X}_1=7.14-1.61X_2$				第二步:$\hat{Y}=4.09-0.41X_2$			
X_1	X_2	$\hat{X}_1$	$X_1-\hat{X}_1$	Y	X_2	$\hat{Y}$	$Y-\hat{Y}$
0	4	0.70	−0.70	1	4	2.45	−1.45
1	4	0.70	0.30	3	4	2.45	0.55
3	1	5.53	−2.53	2	1	3.68	−1.68
6	2	3.92	2.08	6	2	3.27	2.73
8	0	7.14	0.86	4	0	4.09	−0.09
第 3 步:			$r_{(x_1-\hat{x}_1)(y-\hat{y})}$		=	$r_{y1 \cdot 2}=0.89$	

当控制 X_1 来计算 Y 与 X_2 的偏相关时（请读者来练习），我们能清楚地看到这一原理的强烈反差。原始相关是 $r_{y2}=-0.38$，而偏相关是 $r_{y2\cdot1}=0.77$，这意味着结果已经改变。学习成绩与学生声望之间的最初关系是中等并且是负向的（成绩越好，名声越差），但对于课外活动处于相同水平的学生来说，二者关系则很强并且是正向的（成绩越好，名声越好）。这种相关与偏相关之间的巨大差别，原因是控制变量 X_1 与 X_2 和 Y 均有很强的关联（$r_{12}=-0.86$，$r_{1y}=0.75$）。

偏相关系数的计算

容易推出，偏相关系数（$r_{y1\cdot2}$）等于一个分数，分子是原始相关（r_{y1}）与边缘关系乘积（$r_{y2}r_{12}$）的差值，分母是原始变量 Y 和 X_1 不能被控制变量 X_2 解释的方差比例（即剩余方差 $1-r_{y2}^2$ 和 $1-r_{12}^2$）之乘积的几何均值（乘积开根号）：

$$r_{y1\cdot2}=\frac{r_{y1}-r_{y2}r_{12}}{\sqrt{[(1-r_{y2}^2)(1-r_{12}^2)]}}$$

我们可以用很多方法得到这个公式。肯德尔和斯图尔特的矩阵符号方法（Kendall & Stuart, 1969, pp. 317-318）和布朗利更详细的代数算记法（Brownlee, 1965, pp. 429-430）中都有这个推导。肯德尔和斯图尔特（Kendall & Stuart, 1969, pp.328-329）还成功地在矢量模型中用几何学推导出该公式，对于喜欢智力艺术的读者，我们诚挚地推荐这一方法。另外一种计算 $r_{y1\cdot2}$ 公式的方式是路径分析【Path Analysis】。路径系数是标准化的偏回归系数，可以表示为零阶相关系数的函数，如下所示：

$$b_{y1\cdot2}^{*}=\frac{r_{y1}-r_{y2}r_{12}}{1-r_{12}^2}$$

当 X_1 为因变量，Y 为自变量，则表示为：

$$b_{1y\cdot2}^{*}=\frac{r_{y1}-r_{y2}r_{12}}{1-r_{y2}^2}$$

正如零阶相关系数 r_{y1} 是两个回归系数 b_{y1} 和 b_{1y} 的几何均值，一阶偏相关系数 $r_{y1\cdot2}$ 是两个偏回归系数 $b_{y1\cdot2}$ 和 $b_{1y\cdot2}$（或表示为它们的标准化形式 $b_{y1\cdot2}^{*}$ 和 $b_{1y\cdot2}^{*}$）的几何均值，如下所示：

$$b_{y1\cdot2}^{*}b_{1y\cdot2}^{*}=\left[\frac{(r_{y1}-r_{y2}r_{12})}{(1-r_{12}^2)}\frac{(r_{y1}-r_{y2}r_{12})}{(1-r_{y2}^2)}\right]^{\frac{1}{2}}$$

$$=\frac{r_{y1}-r_{y2}r_{12}}{[(1-r_{12}^2)(1-r_{y2}^2)]^{\frac{1}{2}}}=r_{y1\cdot2}$$

该公式可扩展到高阶系数。例如，如果我们探索 X_1 与 Y 之间的关系，在控制变量 X_2 之外，再引入控制变量 X_3，那么可以先直接写出 $r_{y1\cdot2}$ 的公式，然后在公式中每个零阶相关中加入控制变量 X_3。或者颠倒顺序，我们可以先写出 $r_{y1\cdot3}$ 的公式，然后在每一处添加控制变量 X_2：

$$r_{y1\cdot23}=\frac{r_{y1\cdot3}-r_{y2\cdot3}r_{12\cdot3}}{[(1-r_{12\cdot3}^2)(1-r_{y2\cdot3}^2)]^{\frac{1}{2}}}$$

$$=\frac{r_{y1\cdot2}-r_{y3\cdot2}r_{13\cdot2}}{[(1-r_{13\cdot2}^2)(1-r_{y3\cdot2}^2)]^{\frac{1}{2}}}$$

可以看出，二阶偏相关系数是一阶系数的一个函数。每一个高阶系数都可以通过这一方式（即根据它的低阶系数）写出。

显著性检验

偏相关系数和相应的偏回归系数的显著性检验是相同的。详细讨论参见回归分析【Regression Analysis】。

——Jacques Tacq

（叶鹏飞译　高勇校）

参考文献

Blalock, H. M. (1972). *Social statistics.* Tokyo: McGraw-Hill Kogakusha.

Brownlee, K. (1965). *Statistical theory and methodology in science and engineering.* New York: John Wiley.

Hays, W.L. (1972). *Statistics for the social sciences.* New York: Holt.

Kendall, M., & Stuart, A. (1969). *The advanced theory of statistics: Vol. 2. Inference and relationship.* London: Griffin.

Tacq, J. J. A. (1997). *Multivariate analysis techniques in social science research: From problem to analysis.* London: Sage.

偏最小二乘回归(Partial Least Squares Regression)

偏最小二乘(PLS)回归是根据和结合了主成分分析【Principal Component Analysis】和多元回归【Multiple Regression】的一项新的方法。当我们需要根据大量的自变量(即预测因子)来预测一系列的因变量时,这种方法是非常有效的。它起源于社会科学(特别是经济学;Wold, 1966),但首次在化学计量学(部分归功于沃德的儿子斯万特)(参见,如 Geladi & Kowalski, 1986)和感官评价(Martens & Naes, 1989)中得以大量使用。但是 PLS 回归也成了社会科学中实验和非实验数据的多元计算方法(如神经影像学;参见 McIntosh, Bookstein, Haxby, & Grady, 1996)。它最早是作为类似一种幂方法(用于计算特征向量)的算法提出来的,但很快得到一种统计框架下的解释(Frank & Friedman, 1993; Helland, 1990; Höskuldsson, 1988; Tenenhaus, 1998)。

先决条件的概念和符号

由 K 个因变量描述的 I 个样本被写成 $I \times K$ 矩阵记作 **Y**,I 个样本的 J 个预测因子的值被汇集在 $I \times J$ 矩阵 **X** 中。

目标

PLS 回归的目标是从 **X** 中预测 **Y** 值,并描述它们共同的结构。当 **Y** 是一个向量,**X** 是满秩,这个目标可以通过最小二乘【Ordinary Least Squares】来完成。当预测因子的数量比观测样本的数量大时,**X** 可能是奇数的,回归方法就不再可行(原因是多重共线性【Multicollinearity】)。几种方法可以解决这一问题。一种方法是剔除一些预测因子(例如采用逐步【Stepwise】法);另一种方法称为主成分回归,是进行 **X** 矩阵上的主成分分析【Principal Component Analysis, PCA】,然后将 **X** 的主成分作为 **Y** 的回归变量。

主成分的正交性消除了这些多重共线性问题。但选择预测因子最佳子集的问题仍然存在。一种可能的策略是只保留前几个主成分。但它们是被选择用来解释 **X** 而不是 **Y** 的,因此无法保证这些“解释”**X** 的主成分是与 **Y** 相关的。

相比之下,PLS 回归从 **X** 中发现那些同样与 **Y** 相关的成分。特别是 PLS 回归探寻一组成分[称为潜在向量(latent vectors)],这些成分同时分解 **X** 和 **Y** 的约束,能够尽可能地解释 **X** 与 **Y** 之间的协方差。这一步

骤推广了 PCA。其后紧接着一个回归过程，用 $\mathbf{X}$ 的分解来预测 $\mathbf{Y}$。

预测因子和因变量的同时分解

PLS 回归将 $\mathbf{X}$ 和 $\mathbf{Y}$ 分解为同一组正交因子与其特定载荷的乘积。所以，自变量分解为 $\mathbf{X}=\mathbf{TP}^T$，其中 $\mathbf{T}^T\mathbf{T}=\mathbf{I}$，$\mathbf{I}$ 是单位矩阵（在该方法的某些衍生方法中，$\mathbf{T}$ 不需要具有单位范数）。以此类推，对于 PCA，$\mathbf{T}$ 称为得分矩阵，$\mathbf{P}$ 为载荷矩阵（在 PLS 回归中，载荷不是正交的）。同样，$\mathbf{Y}$ 被估计为 $\hat{\mathbf{Y}}=\mathbf{TBC}^T$，其中 $\mathbf{B}$ 是一个"回归权重"为对角元素的对角矩阵（更多权重的细节请见下文）。$\mathbf{T}$ 的列是潜在向量。当其数量等于 $\mathbf{X}$ 的秩时，它们进行的是对 $\mathbf{X}$ 的精确分解。需要注意的是，它们仅仅是估计 $\mathbf{Y}$（即一般来讲，$\hat{\mathbf{Y}}$ 不等于 $\mathbf{Y}$）。

PLS 回归和协方差

潜在向量的选择方法有很多。实际上，在先前的公式中，任何一组能够生成 $\mathbf{X}$ 列空间的正交向量都可被用来充当 $\mathbf{Y}$。为了得到 $\mathbf{T}$，还需要其他条件。对于 PLS 回归，这相当于寻找两组权重 $\mathbf{w}$ 和 $\mathbf{c}$ 来建立（分别）$\mathbf{X}$ 和 $\mathbf{Y}$ 列的线性组合，使它们的协方差最大。具体地说，目标是要在约束条件 $\mathbf{w}^T\mathbf{w}=1$，$\mathbf{t}^T\mathbf{t}=1$ 和 $\mathbf{t}^T\mathbf{u}$ 最大的情况下，获得第一对向量 $\mathbf{t}=\mathbf{Xw}$ 和 $\mathbf{u}=\mathbf{Yc}$。当找到第一个潜在向量时，它被同时从 $\mathbf{X}$ 和 $\mathbf{Y}$ 中剥离出来（这保证了潜在向量的正交性），并且该过程一直持续到 $\mathbf{X}$ 变成一个零矩阵（细节请见算法部分）。

一种 PLS 回归的算法

PLS 回归的性质可以从原来的算法中得到。第一步是建立两个矩阵：$\mathbf{E}=\mathbf{X}$ 和 $\mathbf{F}=\mathbf{Y}$。然后将这两个矩阵进行列中心化和标准化（即转化为 Z 分数）。这些矩阵的平方和表示为 SS_X 和 SS_Y。在开始迭代处理之前，向量 $\mathbf{u}$ 用随机值进行初始化（在下文中，符号 $\propto$ 表示"对运算结果进行标准化"）。

第一步：$\mathbf{w}\propto\mathbf{E}^T\mathbf{u}$（估计 $\mathbf{X}$ 权重）；

第二步：$\mathbf{t}\propto\mathbf{Ew}$（估计 $\mathbf{X}$ 因子得分）；

第三步：$\mathbf{c}\propto\mathbf{F}^T\mathbf{t}$（估计 $\mathbf{Y}$ 权重）；

第四步：$\mathbf{u}=\mathbf{Fc}$（估计 $\mathbf{Y}$ 得分）。

如果 $\mathbf{t}$ 不收敛，就转到第一步；如果 $\mathbf{t}$ 收敛，则计算 b 值，$b=\mathbf{t}^T\mathbf{u}$，用于通过 $\mathbf{t}$ 预测 $\mathbf{Y}$ 值，并计算 $\mathbf{X}$ 上的因子载荷 $\mathbf{p}=\mathbf{E}^T\mathbf{t}$。接着将 $\mathbf{t}$ 的效应从 $\mathbf{E}$ 和 $\mathbf{F}$ 中剥离出来：$\mathbf{E}=\mathbf{E}-\mathbf{tp}^T$ 和 $\mathbf{F}=\mathbf{F}-b\mathbf{tc}^T$。然后将向量 $\mathbf{t}$，$\mathbf{u}$，$\mathbf{w}$，$\mathbf{c}$ 和 $\mathbf{p}$ 存放到相应的矩阵，标量 b 存储为 $\mathbf{B}$ 的对角元素。被潜在向量解释的 $\mathbf{X}$（或者对于 $\mathbf{Y}$）的平方和计算为 $\mathbf{p}^T\mathbf{p}$（对应 Y 的 b^2），解释的方差比例则通过解释的方差除以相应的总平方和（即 SS_X 和 SS_Y）来计算。

如果 $\mathbf{E}$ 是一个零矩阵，那么就找到了整个潜在向量的集合；否则，从步骤 1 重新开始计算过程。

PLS 回归和奇异值分解

上述迭代算法与幂方法相似，也用来寻找特征向量。所以 PLS 回归与特征值和奇异值分解密切相关。例如，如果我们从步骤 1 开始，计算 $\mathbf{w}\propto\mathbf{E}^T\mathbf{u}$，反复替换最右边的项，我们可以得到一连串等式：$\mathbf{w}\propto\mathbf{E}^T\mathbf{u}\propto\mathbf{E}^T\mathbf{Fc}\propto\mathbf{E}^T\mathbf{FF}^T\mathbf{t}\propto\mathbf{E}^T\mathbf{FF}^T\mathbf{Ew}$。这表明第一权重向量 $\mathbf{w}$ 是矩阵 $\mathbf{X}^T\mathbf{Y}$ 的第一右奇异向量。同样的，第一权重向量 $\mathbf{c}$ 是矩阵 $\mathbf{X}^T\mathbf{Y}$ 的第一左奇异向量。同理可以得到，第一向量 $\mathbf{t}$ 和 $\mathbf{u}$ 是 $\mathbf{XX}^T\mathbf{YY}^T$ 和 $\mathbf{YY}^T\mathbf{XX}^T$ 的第一特征向量。

因变量的预测

因变量使用多元回归方程进行预测：$\hat{\mathbf{Y}}=\mathbf{TBC}^T=\mathbf{XB}_{PLS}$，其中 $\mathbf{B}_{PLS}=(\mathbf{P}^{T+})\mathbf{BC}^T$

($\mathbf{P}^{T+}$是$\mathbf{P}^{T}$的 Moore-Penrose 伪逆)。如果使用了所有 **X** 的潜在变量,该回归等价于主成分回归。当仅仅使用了潜在变量的一个子集时,这个特定数量的预测因子对 **Y** 的预测是最佳的。

一个明显的问题是,找到能使新观察值的预测实现最佳一般化所需要的潜在变量个数。这通常可以通过交叉验证方法比如自举法【Bootstrapping】来实现。

潜在变量的解释通常通过绘图来进行,这种图类似于 PCA 表(即将观察值绘制在$\mathbf{t}_1 \times \mathbf{t}_2$的空间中)。

一个实例

我们想预测对一组五瓶葡萄酒的主观评价。要预测的因变量是每种酒的受欢迎程度,以及它搭配肉类与甜点的口感(由一个专家小组认定)(表 1)。预测因子是每一种酒的价格、糖含量、酒精度和酒的酸度(表 2)。

表 1 因变量的 Y 矩阵

酒	受欢迎度	与肉类搭配	与甜点搭配
1	14	7	8
2	10	7	6
3	8	5	5
4	2	4	7
5	6	2	4

表 2 预测因子的 X 矩阵

酒	价 格	糖含量	酒精度	酸 度
1	7	7	13	7
2	4	3	14	7
3	10	5	12	5
4	16	7	11	3
5	13	3	10	3

PLS 回归建立的不同矩阵见表 3 至表 11。从表 11 中我们可以看出两个潜在向量解释了 **X** 的 98%的方差和 **Y** 的 85%的方差。这表示最终方案是保留这两个维度。二维回归系数(即$\mathbf{B}_{PLS}$,见表 8)的检查,表明糖含量主要影响如何选择甜点酒,而价格与酒的受欢迎程度呈现负相关(至少在这个例子中是这样)。潜在向量显示,$\mathbf{t}_1$表示的是价格,$\mathbf{t}_2$反映的是糖含量。

表 3 矩阵 T

酒	$\mathbf{t}_1$	$\mathbf{t}_2$	$\mathbf{t}_3$
1	0.453 8	−0.466 2	0.571 6
2	0.539 9	0.494 0	−0.463 1
3	0	0	0
4	−0.430 4	−0.532 7	−0.530 1
5	−0.563 3	0.504 9	0.421 7

表 4 矩阵 U

	$\mathbf{u}_1$	$\mathbf{u}_2$	$\mathbf{u}_3$
1	1.945 1	−0.761 1	0.619 1
2	0.934 7	0.530 5	−0.538 8
3	−0.232 7	0.608 4	0.082 3
4	−0.915 8	−1.157 5	−0.613 9
5	−1.731 3	0.779 7	0.451 3

表 5 矩阵 P

	$\mathbf{p}_1$	$\mathbf{p}_2$	$\mathbf{p}_3$
价格	−1.870 6	−0.684 5	−0.179 6

续表

	$\mathbf{p}_1$	$\mathbf{p}_2$	$\mathbf{p}_3$
糖度	0.046 8	−1.997 7	0.082 9
酒精度	1.954 7	0.028 3	−0.422 4
酸度	1.987 4	0.055 6	0.217 0

表 6　矩阵 **W**

	$\mathbf{w}_1$	$\mathbf{w}_2$	$\mathbf{w}_3$
价格	−0.513 7	−0.337 9	−0.349 2
糖度	0.201 0	−0.940 0	0.161 2
酒精度	0.570 5	−0.018 8	−0.821 1
酸度	0.608 5	0.042 9	0.421 8

表 7　当三个潜在向量被使用时的矩阵 $\mathbf{B}_{PLS}$

	受欢迎度	与肉类搭配	与甜点搭配
价格	−1.060 7	−0.074 5	0.125 0
糖度	0.335 4	0.259 3	0.751 0
酒精度	−1.414 2	0.745 4	0.500 0
酸度	1.229 8	0.165 0	0.118 6

表 8　当两个潜在向量被使用时的矩阵 $\mathbf{B}_{PLS}$

	受欢迎度	与肉类搭配	与甜点搭配
价格	−0.266 2	−0.249 8	0.012 1
糖度	0.061 6	0.319 7	0.790 0
酒精度	0.296 9	0.367 9	0.256 8
酸度	0.301 1	0.369 9	0.250 6

表 9　矩阵 **C**

	$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{c}_3$
受欢迎度	0.609 3	0.051 8	0.967 2
与肉类搭配	0.702 4	−0.268 4	−0.218 1
与甜点搭配	0.368 0	−0.961 9	−0.130 1

表 10　向量 **b**

$\mathbf{b}_1$	$\mathbf{b}_2$	$\mathbf{b}_3$
2.756 8	1.627 2	1.119 1

表 11　用潜在向量解释的 **X** 与 **Y** 的方差

潜在向量	解释 **X** 方差的%	解释 **X** 方差的累积%	解释 **Y** 方差的%	解释 **Y** 方差的累积%
1	70	70	63	63
2	28	98	22	85
3	2	100	10	95

与其他方法的关系

PLS 回归明显与典型相关【Canonical Correlation】和多元因子分析【Factor Analysis】有关(Escofier & Pagès, 1988)。特嫩豪斯(Tenenhaus,1998)、帕热斯和特嫩豪斯(Pagès & Tenenhaus,2001)详细探讨了这些关系。PLS 回归的主要创造性在于保留了预测因子和因变量之间关系的不对称性,而这些其他方法将其看成是对称的。

软件

PLS 回归需要复杂的计算,它的应用依赖于可利用的软件。在化学中有两个主要

软件:第一个叫 SIMCA-P,由沃尔德最初开发;第二个叫 UNSCRAMBLER,是由马滕斯首先开发的,他是这个领域的另一位先驱人物。在脑成像领域,SPM 是应用最广泛的软件之一,最近(2002)集成了 PLS 回归模块。这些领域之外,SAS PROC PLS 可能是最容易获得的程序。另外,感兴趣的读者可以从作者主页下载一组 MATLAB 程序。公共版的 MATLAB 程序及其教程和案例,可以从 N-Way 项目的主页获得。最后,EIGENRESEARCH 公司也开发了商用 MATLAB 程序包。

——Hervé Abdi

(叶鹏飞译 高勇校)

参考文献

Escofier, B., & Pagès, J. (1988). *Analyses factorielles multiples* [Multiple factor analyses]. Paris: Dunod.

Frank, I. E., & Friedman, J. H. (1993). A statistical view of chemometrics regression tools. *Technometrics*, *35*, 109-148.

Geladi, P., & Kowalski, B. (1986). Partial least square regression: A tutorial. *Analytica Chemica Acta*, *35*, 1-17.

Helland, I. S. (1990). PLS regression and statistical models. *Scandinavian Journal of Statistics*, *17*, 97-114.

Höskuldsson, A. (1988). PLS regression methods. *Journal of Chemometrics*, *2*, 211-228.

Martens, H., & Naes, T. (1989). *Multivariate calibration*. London: Wiley.

McIntosh, A. R., Bookstein, F. L., Haxby, J. V., & Grady, C. L. (1996). Spatial pattern analysis of functional brain images using partial least squares. *Neuroimage*, *3*, 143-157.

Pagès, J., & Tenenhaus, M. (2001). Multiple factor analysis combined with PLS path modeling: Application to the analysis of relationships between physicochemical variables, sensory profiles and hedonic judgments. *Chemometrics and Intelligent Laboratory Systems*, *58*, 261-273.

Tenenhaus, M. (1998). *La régression PLS* [PLS regression]. Paris: Technip.

Wold, H. (1966). Estimation of principal components and related models by iterative least squares. In P. R. Krishnaiaah (Ed.), *Multivariate analysis* (pp. 391-420). New York: Academic Press.

偏回归系数(Partial Regression Coefficient)

偏回归系数,又称回归系数【Regression Coefficient】、回归权重、偏回归权重、斜率系数或偏斜率系数。它用于多元线性回归(MLR)分析中,是当一个自变量【Independent Variable】增加一个单位,同时其他自变量保持不变时,因变量【Dependent Variable】所增加的量。此系数之所以称为偏系数,是因为它的数值通常取决于其他自变量。特别是一个自变量的偏系数的数值变化,取决于回归方程式中其他自变量。

多元回归框架

多元线性回归的目的,是从 N 个题目收集的测量值中,通过一组 K 个自变量 $\{X_1, \cdots, X_k, \cdots, X_K\}$ 来预测【Predict】因变量 Y 的值。我们用 $\mathbf{X}$ 表示 $N\times(K+1)$ 增强矩阵【Matrix】,收集自变量数据(此矩阵之所以称为增强,是因为第一列仅仅由 1 组成),同时用 $\mathbf{y}$ 表示因变量的 $N\times1$ 观察向量

(图 1):

$$\mathbf{X}=\begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} & \cdots & x_{1K} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} & \cdots & x_{nK} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & x_{N1} & \cdots & x_{Nk} & \cdots & x_{NK} \end{pmatrix}$$

$$\mathbf{y}=\begin{pmatrix} y_1 \\ \vdots \\ y_n \\ \vdots \\ y_N \end{pmatrix}$$

图 1 X 和 y 矩阵的结构

多元回归有一系列偏回归系数 b_k,这样,因变量就能够尽可能地接近自变量的线性组合(b_k成为组合的权重)。因此,因变量的预测值 $\hat{Y}$ 的获得方式为:

$$\hat{Y}=b_0+b_1X_1+b_2X_2+\cdots+ b_kX_k+\cdots+b_KX_K \qquad (1)$$

偏回归系数的数值计算采用最小二乘【Ordinary Least Squares, OLS】。通常比较方便地用矩阵符号来表示 MLR 等式。在这个框架中,因变量的预测值 $\hat{Y}$ 被收集在向量 $\hat{y}$ 中,并且通过 MLR 得到:

$$\hat{\mathbf{y}}=\mathbf{Xb},\mathbf{b}=(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \qquad (2)$$

预测值的质量是采用多元相关系数的平方 $R^2_{Y,1,\cdots,k,\cdots,K}$来进行评估,即因变量($Y$)和预测的因变量($\hat{Y}$)之间相关系数的平方。每个自变量对于回归方程的贡献是通过与每个自变量有关的偏相关系数进行评估的。与偏回归系数密切联系,这一系数与在包含其他所有自变量后,在回归方程中加入此变量所获得的解释方差的增加量是一致的。

偏回归系数和回归系数

当自变量两两正交时,每一个自变量在回归中的效应是通过计算自变量和因变量之间的回归斜率来估算的。在这种情况下(也就是自变量之间的正交性),偏回归系数等同于回归系数。在所有其他情况下,回归系数将不同于偏回归系数。

表 1 一组数据:从 X_1 和 X_2 来预测 Y

Y(记忆广度)	14	23	30	50	39	67
X_1(年龄)	4	4	7	7	10	10
X_2(语速)	1	2	2	4	3	6

来源:Abdi, Dowling, Valentin, Edelman, & Posamentier (2002).

注意:Y 是一个儿童短时间内能够记住的数字数量("记忆广度"),X_1 是儿童的年龄,X_2 是儿童的语速(在一个规定时间内该儿童能够拼读多少单词)。6 个儿童参加测试。

例如,考虑表 1 给出的数据,通过自变量 X_1 和 X_2 预测因变量 Y。在这个案例中,Y 是一个儿童的记忆广度(即一个孩子在短时间记忆的任务中能记住的单词数量),我们希望通过儿童的年龄(X_1)和语速(X_2)来进行预测。预测方程[采用方程(2)]为 $\hat{Y}=1.67+X_1+9.50X_2$,其中 b_1 等于 1,b_2 等于 9.50。这意味着儿童记忆广度的增加,是每年一个单词和每拼读一个额外单词就增加 9.50个单词(即他们说得越快,能记忆的单词也越多)。

这个数据集的多元线性回归分析产生了一个多元相关系数的平方 $R^2_{Y\cdot1,2}=0.9866$。X_1 和 X_2 之间的相关系数为 $r_{1,2}=0.7500$,X_1 和 Y 之间的相关系数为 $r_{Y,1}=0.8028$,X_2 和 Y 之间的相关系数 $r_{Y,2}=0.9890$,X_1 和 Y 之间偏回归系数的平方计算为 $r^2_{Y\cdot1|2}=R^2_{Y\cdot12}-r^2_{Y,2}=0.9866-0.9890^2=0.0085$;同样的,$r^2_{Y\cdot2|1}=R^2_{Y\cdot12}-r^2_{Y,1}=$

$0.986\,6-0.802\,8^2=0.342\,1$。

偏回归系数的 F 和 t 检验

偏回归系数为 0 的虚无假设【Null Hypothesis】可以使用标准的 F 检验【F Test】,它原是用来检验相应的偏相关系数为 0 的虚无假设。这个 F 检验的自由度为 $\nu_1=1,\nu_2=N-K-1$(N 为观察的数量,K 为预测因子的数量)。因为自由度 ν_1 等于 1,F 的平方根给出一个学生 t 检验。F 的计算最好用一个例子描述,在我们的案例中,变量 X_1 的 F 值公式如下:

$$F_{Y\cdot 1,2}=\frac{r^2_{Y\cdot 1|2}}{1-R^2_{Y\cdot 1,2}}\times df_{回归}$$

$$=\frac{r^2_{Y\cdot 1|2}}{1-R^2_{Y\cdot 1,2}}\times(N-K-1)$$

$$=\frac{0.008\,5}{1-0.986\,6}\times 3=1.91$$

对于给定的显著性水平 0.05,当自由度 $\nu_1=1,\nu_2=N-K-1=3$ 时,F 值小于临界值 10.13。因此,b_1(以及 $r^2_{Y\cdot 1|2}$)不能被认为不等于 0。

标准误和置信区间

偏回归系数的标准误【Standard Error】对于置信区间【Confidence Interval】的计算以及对于实施其他统计检验是有用的。系数 b_k 的标准误表示为 S_{b_k}。它可以直接通过 F 值计算,$S_{b_k}=b_k\times\sqrt{1/F_k}$,其中 F_k 是 b_k 的 F 检验值。例如,我们得到第一个变量的值:$S_{b_k}=1\times\sqrt{1/1.91}=0.72$。

b_j 的置信区间为 $b_k\pm S_{b_k}\sqrt{F_{临界值}}$,$F_{临界值}$ 是 F 检验的临界值。例如,b_1 的置信区间等于 $1\pm0.72\times\sqrt{10.13}=1\pm2.29$,因此 b_1 的 95% 置信区间是从 $-1.29\sim+3.29$。这个区间包含了 0 值,也表示不能拒绝虚无假设。

β 权重和偏回归系数

这里有一些困惑,因为根据此文,β 是通过 b 值来估计的参数,然而在其他文章中,β 是当回归中所有变量以 Z 得分(Z-scores)形式表示时得到的回归权重。在后一种情况中,β 被称为标准化偏回归系数或 β 权重。这些权重有一个优势,就是一个自变量的权重与另一个自变量的权重是可以相互比较的,原因是测量单位的影响已经被排除。β 权重可以很容易地用 b 计算出来,$\beta_k=\frac{S_k}{S_y}b_k$($S_k$ 为第 k 个自变量的标准偏差,S_y 为因变量的标准偏差)。

——Hervé Abdi

(叶鹏飞译　高勇校)

参考文献

Abdi, H., Dowling, W. J., Valentin, D., Edelman, B., & Posamentier, M. (2002). *Experimental design and research methods.* Unpublished manuscript, University of Texas at Dallas, Program in Cognition and Neuroscience.

Cohen, J., & Cohen, P. (1984). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Darlington, R. B. (1990). *Regression and linear models.* New York: McGraw-Hill.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis.* New York: John Wiley.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research* (3rd ed.). New York: Wadsworth.

参与观察(Participant Observation)

参与观察是这样一种资料搜集方法,研究者在一个持续性的区域社会生活中通过参与来观察它。研究者之前可能是也可能不是这个所研究社会环境的自然成员。参与者常用的经典形式是成为一个小规模面对面团体中相对长期的成员,但是也可能延伸到匿名加入某个群体;因此参与者和非参与者之间的分界线并不明显。雷蒙德·戈尔德(Gold,1969)对于几种可能的田野角色的区分——从完全的参与者到作为观察者的参与者/作为参与者的观察者/完全的观察者——已经广泛地被用来说明可能的范围。参与观察者的实际行为可以变化很大,有时可能会选择对他们各自合适的相互分离的方法,比如无结构式访谈【Unstructured Interview】、对物理环境特征的观察或书面材料的收集。

将参与观察归类为定性方法已经成为惯例,但是它的数据有时候是可以量化的,尽管观察环境经常会为标准化材料的收集带来困难。同样,它通常假设观察者不会去检验假设,而是以开放的态度进入田野去观察生活的面貌。这种数据收集模式并不妨碍假设的检验,但是如果对某一情境及其中成员的意义不能足够熟悉,就无法建立适当的假设。参与观察通常用于研究工作环境中的团队、小型社区中的非正式组群或朋友群体的社会生活等;这些都是可以发展面对面互动的情境,而且单独一个研究者或一个小型工作团队容易接近并确立角色。然而,举例来说,参与式观察也用于研究足球爱好者、右翼政治团体和学校。这种方法不太适合大规模总体的研究或预设变量之间因果关系的代表性数据的收集;运用的是个案研究【Case Study】的逻辑而不是随机样本【Random Sampling】。

历史

“参与观察”的实践历史需要与这一术语的使用历史和它在现代意义上的应用区分开来。在该术语出现之前,甚至远在社会科学成为一个学术领域之前,观察者就通过个人参与来报告他们的发现。1924 年,该术语第一次在社会学的文献中使用,用于说明那些参加会议并向调查者陈述情况的被调查者的作用(第一次世界大战期间在特罗布里恩群岛[Trobriand islands]上,马林诺夫斯基创造了这种形式,运用在社会人类学的田野工作中,尽管人类学和社会学的传统在很大程度上是独立的)。1920 年代和 1930 年代在芝加哥大学由罗伯特·帕克领导进行的大量工作,通常被认为是社会学中第一个现代意义上的参与观察范例,尽管他们当时并不这样称呼它。在这些作品中,一些是帕克鼓励他的学生详细记载过去的经验,一些是在他们的日常生活中自然产生,而不是为了研究目的写作的。这些研究利用了多种资料,并没有认为参与才是理解意义的最佳途径。只是在第二次世界大战以后,再次由芝加哥和那里的学者如霍华德·贝克尔带领,相关的方法论著作取得很大发展。贝克尔关注参与观察,将其作为一种主要的数据收集模式。这些著作中提供的理解现实社会行为及其意义的直接移情法,与最近占支配地位的问卷调查形成对比,并且它的资料看上去比那些仅仅描述行为的标准化问题的答案更优越。

现实与伦理问题

参与观察的使用在研究中产生了一些特殊的实践和伦理问题。参与需要获得进入研究场所的方式。与看门人协商,可以解决公开研究中的进入问题。但是与其他资料收集方法相比,研究者在参与观察中要更多依赖于他们的个人特征;男性和女性,或者黑人和白人研究者,并不拥有相同的进入机会,尽管这种传统的边界有时会被超越。隐蔽性地进行研究可以解决进入问题,却提出了严重的伦理问题,此外还可能引发研究者的持续焦虑;然而,还是有一些不能公开进行的重要工作是隐蔽性地完成的。即使当参与观察是公开进行时,一旦研究者成为一个朋友,研究对象也会忘记他们之间的互动将被收集为资料;随后造成的观察者困境是,持续地重申知情同意,会破坏关系的自然性并因此损害到资料收集。这是两个更广泛问题的一部分:如何心平气和地为了研究目的运用自我而没有土著化【Going Native】;以及如何处理这一现实,即在某种程度上,研究者的存在必然会改变研究的情境。抽样问题本身很少会在参与观察中提及,但是要避免过度依赖于某个没有代表性的知情人,同时也要注意在不同时间和地点去观察研究对象互动的各个侧面。记录资料的工作必须在观察完成后及时完成;可以每天晚上通过回忆补写笔记,也可以频繁地去卫生间作笔记,或者使用袖珍计数装置。这些方法的合适程度,取决于当时的社会情境以及寻求的资料性质。

举例

该方法有三个著名的范例,分别是威廉·怀特(Whyte,1943/1955)的《街角社会》(*Street Corner Society*),霍华德·贝克尔、埃弗里特·休斯和安塞而姆·斯特劳斯(Hughes & Strause, 1961)的《白衣男孩》(*Boys in White*, Becker),以及劳德·汉弗莱斯(Humphreys,1970)的《茶室交易》(*Tearoom Trade*)。怀特对波士顿一个街头帮派的研究非常吸引人,而且因为在后来版本中增加的详细方法附录而更加重要,这促进了此类工作中涉及的实践议题的讨论。《白衣男孩》是一项关于医科学生的研究,其观察和分析方法极其系统,写得也很清晰有条理。《茶室交易》因其在公共卫生间对同性恋行为的观察而声名狼藉,引发伦理问题的是汉弗莱斯识别同性恋参与者的间接方法。

——Jennifer Platt

(叶鹏飞译　高勇校)

参考文献

Becker, H. S., Hughes, E. C., & Strauss, A. L. (1961). *Boys in white*. Chicago: University of Chicago Press.

Gold, R. L. (1969). Roles in sociological field observations. In G. J. McCall & J. L. Simmons (Eds.), *Issues in participant observation* (pp. 30-38). Reading, MA: Addison-Wesley.

Humphreys, L. (1970). *Tearoom trade*. Chicago: Aldine.

Whyte, W. F. (1955). *Street corner society*. Chicago: University of Chicago Press. (Original work published in 1943.)

参与行动研究(Participatory Action Research)

参与行动研究(PAR)涉及地方社区、组织或联盟(在解决直接影响它们的问题中具有合法个人利益)与作为促进者的外部专家之间的合作。他们一起定义问题,学习如何研究它,设计研究方案,分析研究成果,以及设计和实施必要的行动。参与行动研究基于一种信念,相信地方知识的重要性以及所有社会环境中的人们都有能力为了自身的利益采取行动,以使他们生活的环境民主化。

参与行动研究的关键要素包括这些信念:研究的目的是要通过增强人们规划自身未来发展的能力促进民主化的社会变革;内部-外部合作既是可能的也是值得期望的;地方知识对于任何社会变革项目的成功都是必不可少的;以及权力失衡如果不能在某种程度上得到改善,就不会有重大意义的事件发生。参与行动研究还旨在增强地方利益相关者的能力,以便独立承担更多的改革项目,而不是依赖外部专家。

很多思想立场和方法论实践都与参与行动研究有关。有些实践者信奉外部领导的强硬模式,甚至包括挑衅。其他人则认为,如果给予一个受到良好保护的工作空间,人们可以有效地自我组织起来。有些人把激进的社会变革视为参与行动研究的典型特征,但另一些人则认为,任何能够增进福祉、可持续性或公平的变革都是足够的。

参与行动研究的术语众多而复杂。各种各样的行动主义研究彼此之间都存在一些联系(更多内容参见行动研究【Action Research】的词条)。其中包括行为研究、行为科学、行为学习、协同探究、扎根理论【Grounded Theory】的方法论传统以及一般系统理论。这些方法在运用的学科背景上存在较大不同(教育、社会学、社会运动分析、应用人类学、心理学、社会工作)。它们在好成果的定义上存在不同,对变革过程中外部专家作用大小的期望也有所不同。

参与行动研究长期与学术规范和通常的学院派之间关系不和。参与行动研究拒绝思想和行动的分离,而其他学院派的社会科学和人文学科则依赖于这种分离,参与行动研究还从根本上质疑客观性、控制、距离和质量控制等概念。

参与行动研究应用于很多不同的情境——社区、组织、商业单位、志愿者联盟等,因此,单一的案例都无法公正地处理该方法的多样性。启动一个项目有多种方式:被邀请作为一个团体或联盟的外部顾问;成为一个针对特定议题的专题组织成员;被邀请与社区伙伴一起工作;等等。一个项目如何启动非常重要,但是最重要的却不是如何开始,而是它发展出的民主化过程的质量——它的开放性、灵活性与民主目标的一致性。尽管大不相同,在我自己的研究中,有一个简明的例子会有助于阐明这些项目是如何开始和发展的。

在西班牙中南部的一个小镇里,农业经济已变得越来越脆弱。有 300 多个男性每天离开小镇,去马德里从事建筑业工作,数百名女性从事计件缝纫工作。这是一个拥有阶级冲突历史的高度分层的社区,但这座小镇面临着大部分年轻人向其他城市的流失,一些因素影响到了这个极其重视家庭关系的社会当中所有的阶层。为解决这个问题,我们召集了一个包括市政官员、公立和私立学校的教师在内的计划小组,筹备召开

一个40人的行动研究研讨会,这些人是从所有的年龄段、阶层、性别和意识形态中挑选出来的。这个两天的活动,对于社区的历史、理想的未来和可能的未来达成了共识;提出了改善年轻人未来前景的计划;以及创建了志愿行动团队以解决不同的问题。此后几个月该团队聚集在一起精心策划,寻找资源,通过电子邮件与外部促进者联系,与他们一起采取行动,且每3个月都亲自参加会议。

拥有完全不同社会地位、财富和意识形态的人们一起合作的能力经验,每一个都构成了"一块拼图",是对项目的持久性贡献。一些小的行动项目非常成功,但也有一些失败了。现在,那些从初始行动中获得的要素和过程,都已合并到创建人文博物馆、市民中心的公众倡议以及促进旅游业发展的努力当中。

——Davydd J. Greenwood

(叶鹏飞译　高勇校)

参考文献

Fals-Borda, O., & Anisur Rahman, M. (Eds.). (1991). *Action and knowledge: Breaking the monopoly with participatory action-research.* New York: Apex.

Greenwood, D., & Levin, M. (1998). *Introduction to action research: Social research for social change.* Thousand Oaks, CA: Sage.

Horton, M., & Freire, P. (2000). *We make the road by walking* (Bell, B., Gaventa, J., & Peters, J., Eds.). Philadelphia: Temple University Press.

Reason, R., & Bradbury, H. (Eds.). (2001). *The handbook of action research.* London: Sage.

Whyte, W. F. (Ed.). (1991). *Participatory action research.* Newbury Park, CA: Sage.

参与性评估(Participatory Evaluation)

参与性评估是一类项目评估方法,它会积极地将项目工作人员和参与者纳入评估研究的计划和执行中。参与的程度在一个连续体上变动,从分担评估问题和行动的责任,包括数据收集和分析,到完全控制评估过程及其结果,这里评估者充当一名辅导者的角色,以确保工作的质量。一个相关的术语是合作性评估(Collaborative Evaluation),它是参与性评估的一种形式,近于连续体的中心位置;项目参与者承担一种积极的角色,如在设计问题和数据收集工作中,但是项目评估者在这个过程中保持伙伴角色而不是一名辅导者。

所有形式的参与性评估共有的特征是,项目参与者有目的和明确地参与到评估过程中,在一个项目或当地组织或更广泛的社会中引起变化。如果没有专业评估者的协助,很多人也会寻求建立起一个组织或团体从事额外评估的能力。韦弗和库斯坦(Weaver & Cousins, 2001)提供了分析参与性评估研究的五个维度:利益相关者控制评估决策的程度,挑选参与讨论的个人多样性,参与者的权力关系,研究能够实施的容易度,以及人们的参与深度。

参与性评估的起源反映了该术语的一般性质,其根源来自世界各地的不同传统。根据库斯坦和怀廷(Cousins & Whitmore, 1998)的著作,理论家在两类方法之间进行了区分:实践的参与性评估(P-PE)和变化的参与性评估(T-PE)。实践的参与性评估是由这样的方法发展而来,即通过将项目工作人员和参与者纳入评估过程,寻求强化评

估结果的所有权和使用权。迈克尔·奎因·巴顿(Patton,1997)聚焦于应用的评估,突出了"个人因素"的重要性(也就是说,联系那些对评估感兴趣的个人,并向他们提供有用的信息)。组织发展和组织学习指出了不断的循环过程对于提问、收集和分析数据、思考数据以及作出适当改变所具有的价值。在 P-PE 中,评估者能够提供对评估过程的领导,因为他们让其他人参与进来并对其进行指导。

相比之下,在 T-PE 中,项目参与者(至少在理论上)成为评估工作的领导者,通过公民赋权来支持社会变革和社会正义。P-PE 可通过使项目更具作用或更有效率来改善现状(例如,在心理健康项目中改善服务提供),而 T-PE 是一个过程,力求使被剥夺公民权者和受压迫者的生活世界变得更好(例如,减少一个发展中国家农村社区中的贫困)。这一类型包括了过去 50 年间发展起来的众多以公民为基础的评估和研究方法,如包括参与行动研究、参与性农村评价(更名为参与式学习和行动)、参与性监测与评估,以及最近的强能评估。

——Jean A. King

(叶鹏飞译　高勇校)

参考文献

Cousins, J. B. (2003). Utilization effects of participatory evaluation. In T. Kellaghan, D. L. Stufflebeam, & L. A. Wingate (Eds.), *International handbook of educational evaluation* (pp. 245-266). Dordrecht, The Netherlands: Kluwer Academic.

Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions in Evaluation*, *80*, 5-23.

King, J. A. (1998). Making sense of participatory evaluation practice. *New Directions for Evaluation*, *80*, 57-67.

Patton, M. Q. (1997). *Utilization-focused evaluation* (3rd ed.). Thousand Oaks, CA: Sage.

Weaver, L., & Cousins, B. (2001, November). *Unpacking the participatory process*. Paper presented at the annual meeting of the American Evaluation Association, St. Louis, MO.

Pascal 分布(Pascal Distribution)

Pascal 分布用于对独立伯努利【Bernoulli】试验中,获得特定成功次数(r)所需要的试验次数(x)进行模型化。离散型随机变量【Random Variable】X 可以取从 r 到无穷大的任何整数值。如果有幸每次试验结果均是成功的,那么 $X=r$。若使 $X=x$,那么第 x 个试验必须成功,并且之前的 $x-1$ 次试验中恰好$r-1$次成功。这个定义与二项分布【Binomial Distribution】是相反的。也就是说,Pascal 分布中的随机变量是达到一定成功次数时的试验次数;而在二项分布中的随机变量是特定试验次数中出现的成功次数。由于这个原因,Pascal 分布也称为负二项分布【Negative Binomial Distribution】。虽然这两个分布常常同义地使用,但是 Pascal 分布实际上是在随机变量 X 为整数这一特殊情况下的负二项分布。Pascal 分布的一种特殊情况是:当$r=1$时为几何分布,即获得第一次成功时所必需的试验次数。

Pascal 分布的概率质量函数(PMF)为:

$$P(X=x\mid r,p)=\binom{x-1}{r-1}p^{r}(1-p)^{x-r}$$

其中,$x=r,r+1,r+2,\cdots,0\leqslant p\leqslant 1,0<r\leqslant\infty$。

Pascal 分布的随机变量均值为:

$$\mu = \frac{r}{p}$$

以及方差为：

$$\sigma^2 = \frac{r(1-p)}{p^2}$$

Pascal 分布是泊松分布【Poisson Distribution】的一种扩展，即放松了泊松分布参数 λ 的限制。Pascal 分布假定在一个观察时期内，事件的发生率依据伽马分布而变化，而不要求事件的发生率是恒定的。因此，Pascal 分布比泊松分布更灵活，经常用来模型化过度分散的事件计数【Event Count】数据或方差超过均值的数据。

例如，Pascal 分布可用于对通过 10 个议案前必须向立法机构提出的议案数量进行建模。或者，这一试验结果也可用于确定第 r 次成功之前的失败数量。

——Ryan Bakker

（叶鹏飞译　高勇校）

参考文献

Johnson, N. L., & Kotz, S. (1969). *Discrete distributions*. Boston: Houghton-Mifflin.

King, G. (1998). *Unifying political methodology: The likelihood theory of statistical inference*. Ann Arbor: University of Michigan Press.

路径分析(Path Analysis)

尽管“相关关系并不意味着因果关系”这一说法在社会科学的研究方法课中不断被提及，但它的逆命题却是路径分析和结构方程模型【Structural Equation Modeling, SEM】等定量数据分析方法(还包括多元回归分析【Multiple Regression Analysis】、验证性因子分析【Confirmatory Factor Analysis】)的基础。也就是说，在适当的情况下，“因果关系意味着相关关系的存在”。也就是说，如果一个变量和另一个变量有因果关系，那么在统计上和/或实验上控制了所有其他相关变量的情况下，数据中应该能观察到一种相关【Correlation】(或协方差【Covariance】)。这样，正如常在路径图中所见，路径分析可以在一定程度上被描述为这一过程：假设被测量的变量之间符合因果(结构)关系模型，然后检验观察数据的关系符合最初假设的期望关系的程度。

大多数现代路径分析方法都源于生物统计学家赖特·休厄尔的研究(例如，Wright, 1918)，他首次将路径分析应用于分析骨骼测量数据间的相关关系。奥蒂斯·邓肯(Duncan, 1966)首次将这种方法引入社会学研究中，之后在社会科学学科中逐渐受到重视。在 1970 年代到 1980 年代，受约雷斯括格、沃德·基斯林、大卫·威利、达格·瑟尔布姆、彼得·本特勒和本特·米滕等学者在方法论、计算机技术和应用方面所取得的进展的推动，无数应用研究出现在社会和行为科学中。例如，随着观察数据与假设模型符合程度的统计检验方法的发展，研究的焦点从相关转向协方差。此外，将潜在(如未被观察到的)因子引入路径模型的可能性以及更多一般估计技术的发展(如最大似然估计【Maximum Likelihood】、渐进自由分布【Distribution-Free】)，解决了传统相关路径分析的大部分不足。在整个 1980 年代、1990 年代和今天，专业计算机软件的持续改进(如 LISREL, EQS, AMOS, Mplus)，导致了技术复杂性和使用简易性的同时增加，这促进了路径分析的发展，使 SEM 在过去 30 年中得到更多应用。

在本词条中，将通过理论和实例来介绍路径分析的典型步骤。这些步骤包括模型概念化、参数识别和估计、效应分解、数据模型拟合评估、可能的模型修正和交叉验证。关于路径分析和 SEM 的详细说明可参考 Bollen，1989；Kaplan，2000；Kline，1998；Mueller，1996。

基本原则

例如，一个教育研究者可能有一种关于4个概念之间关系的理论：数学自我概念（Mathematics Self-Concept，MSC）、阅读自我概念（Reading Self-Concept，RSC）、任务目标取向（Task Goal Orientation，TGO）和数学能力（Mathematics Proficiency，MP）。前三个概念可通过等级量表工具操作化为得分，而后者则可以操作化为标准化测验得分。使用这些测量变量分别代表各自的概念，该理论可表达为图 1 的形式。每个方框代表一个测量变量，标作 V，单向箭头表示有正式理论证明，箭头尾部的变量是箭头前面的变量的原因（反之不一定成立）。如图 1 所示，假设 RSC 直接影响 TGO，MSC 直接影响 TGO 和 MP，TGO 直接影响 MP。模型中没有原因输入的变量称为自变量或外生变量（MSC 和 RSC），而有原因输入的变量称为因变量或内生变量（TGO 和 MP）。图中还有另外两个单向箭头，一个指向 TGO，另一个指向 MP。这些表示所有无关的剩余因子（“Errors”，记为 E）对每个内生变量的影响。最后，双向箭头表示这些变量除因果关系外，还由于其他原因相互关联（如同时受到模型之外的其他变量的影响）。MSC 和 RSC 之间就假定存在这种关联。需要注意的是，双向箭头并不表示互为因果（即变量之间互相有直接或间接的因果关系）。那属于非递归模型，超出本章讨论范围（参见 Berry，1984；Kline，1998）。

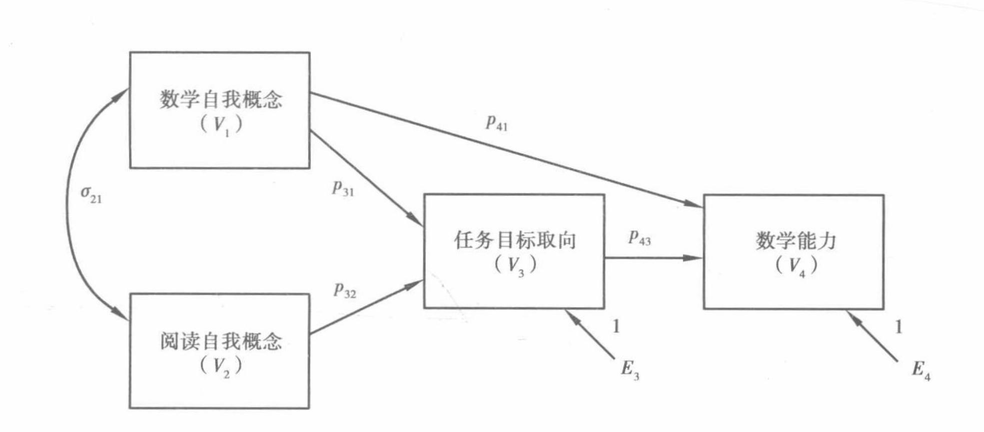


图 1　假设的路径模型

每个单向箭头上标注的是总体中的非标准化路径系数 p，表示相关变量之间因果关系的方向和大小。这类似于，但并不一定等同于非标准化偏回归权重。至于双向箭头，因为它代表两个变量之间的协方差，所以箭头上的符号就是熟悉的 σ。运用这些符号，可以将图 1 中的关系用两种方式表示：结构方程和模型-隐含关系（model-implied relations）。结构方程是以回归方程的形式将内生变量表示为其自变量的函数。假设变量是均值中心化的（由此消除了截距项），当前模型的结构方程表示如下：

$$V_3 = p_{31}V_1 + p_{32}V_2 + 1E_3$$
$$V_4 = p_{41}V_1 + p_{43}V_3 + 1E_4$$

这些方程连同模型中包含的非因果关系(即双箭头),提示我们应当观察数据中的方差和协方差。尤其是,如果总体中这个假设模型成立,那么对此结构方程的期望值【Expected Value】进行代数运算,就可以得到如下总体中的方差(Var)和协方差(Cov):

$$\mathrm{Var}(V_1) = \sigma_1^2$$
$$\mathrm{Var}(V_2) = \sigma_2^2$$
$$\mathrm{Var}(V_3) = p_{31}^2\sigma_1^2 + p_{32}^2\sigma_2^2 + 2p_{31}\sigma_{21}p_{32} + \sigma_{E_3}^2$$
$$\mathrm{Var}(V_4) = p_{41}^2\sigma_1^2 + 2p_{43}p_{31}p_{41}\sigma_1^2 + 2p_{43}p_{32}\sigma_{21}p_{41} + p_{43}^2(p_{31}^2\sigma_1^2 + p_{32}^2\sigma_2^2 + 2p_{31}\sigma_{21}p_{32} + \sigma_{E_3}^2) + \sigma_{E_4}^2$$
$$\mathrm{Cov}(V_1, V_2) = \sigma_{21}$$
$$\mathrm{Cov}(V_1, V_3) = p_{31}\sigma_1^2 + p_{32}\sigma_{21}$$
$$\mathrm{Cov}(V_1, V_4) = p_{41}\sigma_1^2 + p_{31}p_{43}\sigma_1^2 + p_{32}p_{43}\sigma_{21}$$
$$\mathrm{Cov}(V_2, V_3) = p_{32}\sigma_2^2 + p_{31}\sigma_{21}$$
$$\mathrm{Cov}(V_2, V_4) = p_{32}p_{43}\sigma_2^2 + p_{41}\sigma_{21}$$
$$\mathrm{Cov}(V_3, V_4) = p_{43}(p_{31}^2\sigma_1^2 + p_{32}^2\sigma_2^2 + 2p_{31}\sigma_{21}p_{32} + \sigma_{E_3}^2) + p_{31}p_{41}\sigma_1^2 + p_{32}p_{41}\sigma_{21}$$

这 10 个方程式组成了模型的协方差结构,右侧有 9 个未知的总体参数($p_{31}, p_{32}, p_{41}, p_{43}, \sigma_{21}, \sigma_1^2, \sigma_2^2, \sigma_{E_3}^2, \sigma_{E_4}^2$)。如果得到这 4 个变量之间的总体方差和协方差,那么这些数值就可以放在方程左边,然后方程右边的未知参数即可求出(如果这个假设模式是正确的)。如果仅仅知道样本(协)方差,例如表 1 中列出的 1 000 个九年级学生的假设样本数据,它们同样可以放在方程左侧,然后通过推导得出右侧参数的估计值。为了实现这一点,模型中的每个参数都必须可以表达为观察变量(协)方差的函数形式。当只有唯一一种关系式能求得未知参数时,该模型是恰好识别(just-identified)的。当对于一个或多个参数来说有多种表达方式,该模型是过度识别(overidentified)的;在这种情况下,可以导出每个参数估计的最佳拟合(虽然不是唯一的)。然而,如果至少一个参数不能被表示为观察变量(协)方差的函数,该模型就是不足识别的(underidentified),一些参数或全部参数无法仅靠数据进行估计。不足识别的产生,可能是因为研究者试图强加一个相对于观察变量(协)方差个数来说过于复杂的模型(即有太多参数需要估计)。幸运的是,不足识别在大多数路径分析中非常罕见,只在不常见的模式特征中出现(如非递归关系)。

表 1　方差和协方差的样本矩阵

	V_1	V_2	V_3	V_4
V_1(数学自我概念)	1.951			
V_2(阅读自我概念)	−0.308	1.623		
V_3(任务目标取向)	0.262	0.242	1.781	
V_4(数学能力)	28.795	−4.317	17.091	1 375.460

图 1 中的模型是包括 10 个观察(协)方差、9 个待估参数的过度识别模型:1 个协方差,4 个直接路径,2 个变量方差和 2 个误差方差。其自由度为 $10-9=1(df)$。已知一个模型是恰好或(最好是)过度识别的,参数估计可通过由多种不同的估计方法获得。这包括最大似然和广义最小二乘法,二者都假设多元正态,并且是渐近等价和渐近分布无关的估计方法,通常需要大样本量。这些方法迭代地最小化某个差异的函数,该差异是指观察(协)方差和反复用参数估计代替模型-隐含关系而再生的(协)方差之间的差异。在当前的模型中,使用任意 SEM 软件包,可得出关键路径的最大似然估计如下(按照变量的原始单位):$\hat{p}_{31} = 0.163$,$\hat{p}_{32} = 0.180$,$\hat{p}_{41} = 13.742$,$\hat{p}_{43} = 7.575$ 和 $\hat{\sigma}_{21} = -0.308$

（它们的统计显著性均为 0.05 水平）。标准化路径系数，类似于多元回归中的 β 权重，分别是 0.170，0.172，0.518，0.273 和 -0.173。但是，在聚焦这些重要参数之前，我们应该考虑是否有任何的证据表明数据模型（data-model）是不适合的，以及是否有任何统计的（和理论上的证明）原理可以修正假设的模型。

数据模型拟合评估和模型修正

现代路径分析的一个优点是它能评估数据和模型之间的拟合程度。许多方法能帮助研究者决定是否拒绝或者暂时保留一个先验的特定的过度识别模型。一般情况下，由数据观察到的（协）方差和由模型及其参数估计而再生的（协）方差之间拟合度的评估，可分为三类：绝对拟合、简约拟合和增量拟合。绝对拟合（absolute fit）指数越高，观察的（协）方差和根据模型计算出的（协）方差之间的差异就越小，即参数值就越接近于非冗余（协）方差的观察值。这种指数的例子包括：用来检验在总体中没有数据模型不适合这一严格虚无假设的卡方检验统计模型，粗略评估观察（协）方差和再生（协）方差之间平均标准化差异的标准化残差均方根（the standardized root mean square residual，SRMR），以及用于评估能够被模型解释的观察（协）方差信息数量的拟合优度指数（goodness-of-fit index，GFI）。

简约拟合（parsimonious fit）指数考虑的不只是整体的绝对拟合，还要考虑实现这种拟合所要求的模型复杂程度（即参数个数）。诸如此类的指数，如调整拟合优度指数（adjusted goodness-of-fit index，AGFI）、简约拟合优度指数（parsimonious goodness-of-fit index，PGFI）和近似均方根误差（root mean square error of approximation，RMSEA），在模型相对简单（参数较少）且数据拟合较为成功时，取值会更大。最后，增量拟合（incremental fit）指数如赋范拟合指数（normed fit index，NFI）和比较拟合指数（comparative fit index，CFI），是评估一个假设模型相对于只含很少几个参数的基线模型的数据模型拟合程度。

这三种类型的拟合指数共同帮助研究者对路径模型的可接受性进行评估。胡和本特勒（Hu & Bentler，1999）建议对模型的可接受性使用联合标准，如 CFI 值为 0.96 或更大，同时 SRMR 值小于 0.09（或 RMSEA 值小于0.06）。在目前的例子中，不显著的模型卡方值 2.831（自由度为 1），表示如果我们的模型正确地描述了真实的总体关系，那么观察的（协）方差很可能出现。对目前的模型来说这是好消息，但是这个统计量常常会被忽略，因为在大样本的条件下，只要模型中有一点儿误设，它就会变得显著。由此，一般会使用其他拟合指数来进行模型评估。对于目前的模型，当应用上面提出的标准：CFI = 0.997，SRMR = 0.013 及 RMSEA = 0.043，则不存在明显的数据模型不一致。

经过数据模型拟合评估后，必须就该模型的价值得出结论。可接受的拟合指数通常得到的结论是，没有证据表明能够拒绝这个模型或理论。这并不是说，模型和理论已经被证实，更不能说证明它们是正确的；更确切的说法应该是，当前路径模型是能很好地解释观察变量之间关系的众多方法之一。另一方面，当拟合指数表明一个数据模型无法拟合时，人们可能不愿意完全拒绝该模型。相反，通常会尝试着修改模型，通过添加路径，得到可接受的拟合指数。大多数路径分析软件包会通过提供修正指数［拉格朗日乘数检验（Lagrange multiplier tests）］来帮助实现这种模型“改善”，这里的修正指数表明模型中的变化可能在绝对拟合中获得最大的增加，也就是说，减少模型的卡方统计量。尽管这些指数构成了一个潜在的有用工具，能用于补救错误设定的模型，但也

需要警惕得到一个与理论无关却有着最好拟合度的模型。许多替代模型都可以同样很好地解释观察到的数据,因此,模型的修正必须依据对模型背后特定理论的正确理解。此外,如果仅仅在数据模型无法拟合的信息基础上进行修正和再分析,那么随后的拟合结果可能主要是随机造成的,而不是真正来自模型的改进。因此,只要可能,修正后的模型都应该在一个独立样本中进行交叉验证。从目前的分析来看,我们的模型不需要修正指数来提高数据模型拟合的统计显著性(即模型卡方统计量的显著减少);实际上,因为整体模型的卡方值是 2.831(显著性水平>0.05),已经没有提高统计显著性的空间。因此,这里没有要报告的模型修正信息。

最后,给出令人满意的数据模型拟合后,可以得出对于特定模型关系的结论。路径系数的解释类似于回归系数,但通常带有因果倾向。例如,依据假设模型,$\hat{p}_{31}=0.163$ 意味着保持所有其他因素固定不变,MSC 的数值增加 1 个单位,将直接导致 TGO 的数值平均增加 0.163 个单位。回想一下,这是一个统计上显著的影响。或者,可以解释标准化路径系数,如 0.170 的对应值,这表示在假设模型中,保持所有其他因素固定不变时,MSC 数值增加 1 个标准差,直接导致 TGO 数值平均增加 0.170 个标准差。其他非标准化和标准化的直接路径,同样解释了假设模型中一个变量对另一个变量的直接效应(direct effect)。

除了直接效应,也可以得到其他的效应系数。考虑 RSC 和 MP 之间的模型关系。RSC 对 MP 没有直接的影响,但它确实对 TGO 有直接影响,而 TGO 对 MP 有直接影响。这样,RSC 对 MP 就有了间接效应(indirect effect),因为操作性地说,可以追踪相同方向一系列的两个或更多的单箭头来获得一个变量到另一个变量的间接效应。采用非标准化的方法,间接效应的大小是构成它的直接效应的乘积:$(0.180)\times(7.575)\approx 1.364$。在保持所有其他固定不变的条件下,这种非标准化的间接效应值意味着 RSC 数值增加 1 个单位,会间接导致 MP 测试值将平均增加 1.364 个单位。同样,对于标准化方法,标准化间接效应是 $(0.172)\times(0.273)\approx 0.047$,表示在保持所有其他固定不变的条件下,RSC 数值增加 1 个标准差,将间接导致 MP 测试值平均增加 0.047 个标准差。

最后,值得注意的是 MSC 对 MP 既有直接效应也有间接效应。除了非标准化和标准化的直接效应分别是 13.742 和 0.518 之外,非标准化的间接效应可计算为 $(0.163)\times(7.575)\approx 1.235$,标准化的间接效应可计算为 $(0.170)\times(0.273)=0.046$。这意味着 MSC 对 MP 的总的因果影响,即总效应(total effect),是直接效应和间接效应的加总。对于非标准化的方法,总效应是 $13.742+1.235\approx 14.977$,表示在保持所有其他固定不变的条件下,MSC 数值增加 1 个单位,总体上会导致 MP 测试值平均增加14.977个单位。同样的,对于标准化方法,总效应为 $0.518+0.046=0.564$,即保持所有其他固定不变的条件下,MSC 数值增加 1 个标准差,总体上 MP 测试值的标准差将增加0.564个单位。这样,与仅仅提供相关系数和协方差系数相比,通过完整地使用假设模型进行效应分解(effect decomposition),我们可以了解到更多变量之间的关系。

总结

路径分析已成为社会和行为科学众多领域中的一种重要分析工具。它属于结构方程模型技术的体系,允许探究一个先验的理论推导模型中观察变量和潜变量之间的因果关系。路径分析的主要优势在于它能够帮助研究者弥合理论和观察之间常常出

现的差距。正如本词条所强调的,路径分析最好被理解为一个要以理论为引导,以统计为工具的过程。如果理论基础不牢靠,那么其后所有统计学的解释都将是没有说服力的。更多关于路径分析的细节以及 SEM 方法体系的扩展,可以参考的资源如 Bollen, 1989; Kaplan, 2000; Kline, 1998; Mueller, 1996。读者也可能对其他涉及分类变量、以对数线性模型【Log-Linear Model】为基础的类似方法感兴趣(参见 Goodman, 1973; Hagenaars, 1993)。

——Gregory R. Hancock
Ralph O. Mueller
(叶鹏飞译　高勇校)

参考文献

Berry, W. D. (1984). *Nonrecursive causal models*. Beverly Hills, CA: Sage.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley.

Duncan, O. D. (1966). Path analysis: Sociological examples. *American Journal of Sociology*, *72*, 1-16.

Goodman, L. A. (1973). The analysis of multidimensional contingency tables when some variables are posterior to others: A modified path analysis approach. *Biometrika*, *60*, 179-192.

Hagenaars, J. A. (1993). *Loglinear models with latent variables*. Newbury Park, CA: Sage.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1-55.

Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Thousand Oaks, CA: Sage.

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford.

Mueller, R. O. (1996). *Basic principles of structural equation modeling: An introduction to LISREL and EQS*. New York: Springer.

Wright, S. (1918). On the nature of size factors. *Genetics*, *3*, 367-374.

路径系数(Path Coefficient)

参见路径分析【Path Analysis】。

路径依赖(Path Dependence)

路径依赖的概念最初用于经济学,现在在其他社会科学领域出现的频率越来越高,特别是历史社会学和政治学。它是用来解释事件的特定序列如何产生特定结果,以及这种独特序列是如何影响未来选择的。

当一个结果是由于路径依赖产生的,可以观察到一些关于结果的发展过程的特点。第一,存在不可预测性。这意味着,尽管具有对特定期望结果的先验知识,也会发生一些无法预见的事件,它们从历史上看是偶然的。第二,路径依赖强调小事件的重要性。经济学家布瑞恩·亚瑟(Arthur, 1994)指出,“很多结果是可能的,非均匀性、小的不可分性或偶然相遇,会通过将它们加入实际

结果中得到的正反馈而被放大”(p.27)。但最重要的是第三个因子,即“锁定(lock-in)”。当很小的、看似无关紧要的决定沿着一个结果路径产生时,这些决定正在被不断地再造和加强。这是收益递增的逻辑。一旦作出一个决定(例如,关于一种新技术或新制度),最初的决定将很难改变。因为它是被不断使用和改进的,所以它是被“锁定”的。这样,路径依赖过程是无情的:在作出一个决定后,改变的代价非常大,而且随着时间的推移代价会越来越大。这种持久性往往导致出现次优结果,这是路径依赖轨迹的另一个标志。

路径依赖的事例有利于充分理解这个概念。经济学家保罗·大卫(David,1985)在“QWERTY”键盘排列的研究中首次表述了路径依赖。大卫提出,我们现在熟悉的打字机按键顺序是一个随机决定的结果,一旦打字机被按照那种特定的排列制造出来,并且人们开始学习使用它,这种排列就很难再改变。尽管对于QWERTY键盘有很多种可能的替代选择(它们中有一些或许更有效率),但是最初决定使用的那个按键排列随着时间的推移而变成是“锁定”的,并且使用它的人越来越多。结果是改变按键的排列目前在事实上已不太可能。

对路径依赖解释的一个经常性批判是它们简单地宣称历史的重要性。然而,保罗·皮尔逊(Pierson,2000)强调指出,“不是过去本身,而是事件随时间的演变才是理论上最重要的”。(p.264)这种观点得到其他路径依赖支持者的重申,特别是詹姆斯·马哈尼(Mahoney,2000)提出了一个路径依赖分析框架,指出所有路径依赖的论点必须证明:(a)调查的结果对早期发生事件的历史顺序特别敏感;(b)这些早期事件是历史偶然的(不可预测的);(c)这些历史事件必须呈现出实现结果之“路径”的相对惯性或确定性(Mahoney, 2000, pp. 510-511)。这种分析明确地解释了例如特定的制度是如何出现、持续和变得难以改变的。

——Tracy Hoffmann Slagter

(叶鹏飞译 高勇校)

参考文献

Arthur, W. B. (1994). *Increasing returns and path dependence in the economy*. Ann Arbor: University of Michigan Press.

David, P. (1985). Clio and the economics of QWERTY. *American Economic Review*, *75*, 332-337.

Mahoney, J. (2000). Path dependence in historical sociology. *Theory and Society*, *29*, 507-548.

Pierson, P. (2000). Increasing returns, path dependence, and the study of politics. *American Political Science Review*, *94*, 251-267.

路径图(Path Diagram)

参见路径分析【Path Analysis】。

皮尔逊相关系数(Pearson's Correlation Coefficient)

皮尔逊相关系数(记为 r_{xy})是现代用简单相关【Simple Correlation】评估两个变量之间线性【Linear】关系的最佳代表。该系数能表示关系的强度,绝对值为 0~1。数值越大,则变量间的关系越强。系数的符号代表关系的方向,可以是零、正数或负数。变量 X 和 Y 之间的关系为零,表示变量 X 的增加会同时伴有变量 Y 的增加和减少,反之亦然。正相关表示变量 X 的增加(或减少)同时会出现变量 Y 的增加(或减少)。反之,负相关表示变量 X 的增加(或减少)同时会出现变量 Y 的减少(或增加)。

皮尔逊相关系数的起源

在 19 世纪,达尔文的进化论激发了对生物体进行多样化测量的兴趣。高尔顿爵士是当时最著名的贡献者之一,除了遗传定律,他还提出了相关的统计概念。在他 1888 年发表的文章(引自 Stigler,1986)《相关及其测量:基于人种测量学数据》中,高尔顿证明了当两个变量使用相同测量尺度时,前臂和头部长度的回归线具有相同的斜率(记为 r)。他将 r 作为相关(co-relation)的一个指数,并称它为回归系数【Regression Coefficient】。co-relation 的拼写是有意区别于用于其他学科已经使用的 correlation,一年之内在拼写上却又改成了 correlation。然而,皮尔逊相关系数是在皮尔逊使用后才得以为人熟知。

皮尔逊相关系数也被称为皮尔逊积矩相关,定义为:

$$\frac{\sum z_{X_i} z_{Y_i}}{n}$$

它是先将两个变量的 z 分数【z-Scores】相乘(即乘积),然后再取这些乘积的平均值(即矩)。各种皮尔逊相关系数的计算公式都可以从上面的公式中导出(Chen & Popovich,2002),如下所示:

$$\frac{\sum z_{X_i} z_{Y_i}}{n} = \frac{\sum \left(\frac{X_i - \overline{X}}{s_X}\right)\left(\frac{Y_i - \overline{Y}}{s_Y}\right)}{n}$$

$$= \frac{\frac{1}{s_X s_Y} \sum (X_i - \overline{X})(Y_i - \overline{Y})}{n}$$

$$= \frac{\sum \frac{(X_i - \overline{X})(Y_i - \overline{Y})}{n}}{s_X s_Y}$$

$$= \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sqrt{\left(\sum X_i^2 - \frac{\sum X_i^2}{n}\right)\left(\sum Y_i^2 - \frac{(\sum Y_i)^2}{n}\right)}}$$

皮尔逊相关系数的解释

皮尔逊相关系数并不表示两个变量之间的因果关系。它不能被解释为一个比例,也不代表关系成比例的强度。例如,一个 0.4 的相关并不是在强度上等于 0.2 相关的两倍。与皮尔逊相关系数相比,r_{xy}^2(也称为决定系数【Coefficient of Determination】)应当被解释为一个比例。例如,课程成绩和智力之间的皮尔逊相关系数是 0.7。r_{xy}^2 值为0.49,意味着成绩和智力的共享方差是 49%。也可以解释为智力 49%的方差可以由成绩来解释或预测,反之亦然。

皮尔逊相关系数的特殊情况

有几个著名的简单相关系数，如点二列相关系数（the point-biserial correlation coefficient）、Phi 相关系数、斯皮尔曼相关系数【Spearman Correlation Coefficient】、Eta 系数【Eta Coefficient】，它们都是皮尔逊相关系数的特殊情况。假设我们使用皮尔逊相关系数来测量一个二分变量（如性别）和连续型变量（如收入）之间的关系，这种相关通常被称为点二列相关系数。也可用它来测量两个二分变量（如性别和感染）之间的关系，这个系数称为 Phi 相关系数。如果两个变量在定序层次上测量，并且被分配了等级，那么皮尔逊相关系数也可用于这种关系的测量。这个系数经常被称为斯皮尔曼相关系数。最后，当皮尔逊相关系数被用于测量一个多分类变量（如种族、学校、颜色）和一个连续型变量之间的关系时，被称为 Eta 系数（Wherry, 1984）。

皮尔逊相关系数的用途

皮尔逊相关系数能够以多种形式用于多个目标。它可以发挥但不局限于以下功能：(a)作为描述统计量描述两个变量之间的关系；(b)作为一种推断统计【Inferential Statistics】检验总体【Population】中两个变量之间的关系；(c)提供各种信度估计方法，如 Cronbach α【Cronbach's Alpha】、重测信度【Test-Retest Reliability】和折半信度【Split-Half Reliability】；(d)评估效度【Validity】证据；(e)测量干预项目的效应强度。当运用皮尔逊相关系数来对总体 ρ（RHO）进行推断时，二元正态分布【Normal Distribution】的假设是必需的。两个变量 X 和 Y 遵循二元正态分布的充分必要条件是每个可能的线性组合 W 的分布是正态的，$W=c_1X+c_2Y$，其中两个 c 值都不能为零（Hays, 1994）。图 1 所示为二元正态分布当 $\rho=0$ 时的一个例子。

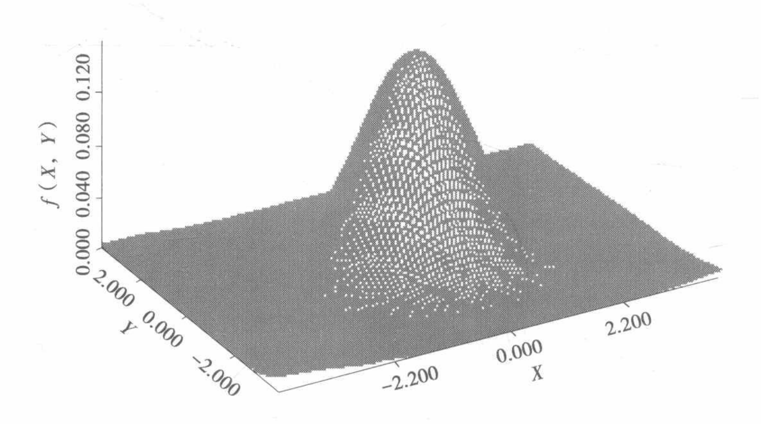


图 1　二元正态分布，设 $\rho=0$ 且 $N=1\ 000$

皮尔逊相关系数的抽样分布

皮尔逊相关系数的抽样分布【Sampling Distribution】的形状是随着总体 ρ 和样本量的函数而变化的，但是当样本量特别大时，其影响是很小的。当总体 ρ 等于零时，其抽样分布是对称的 t 分布【t-Distribution】，如图 2 所示。如下文所述，基于 1 000 个随机产生的样本，当 $\rho=0$ 并且 $n=20$ 时，抽样分布总体上是对称的。因此，为了检验 $\rho=0$ 的虚无假设【Null Hypothesis】，可以进行 t

检验【t-Test】，$t=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$，自由度【Degrees of Freedom】为 $n-2$。

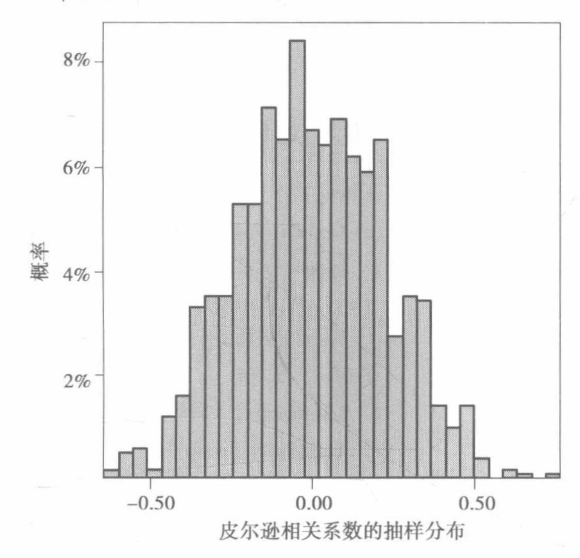


图 2　1 000 个样本的皮尔逊相关系数抽样分布，设 $n=20$ 且 $\rho=0$

然而，对于给定的小样本量，当 $|\rho|$ 增大时，抽样分布会变成负偏斜的【Skewed】，如图 3 所示。如果分布中的每个皮尔逊相关系数都根据 $0.5\times\ln\left(\frac{1+r}{1-r}\right)$ 转换【Transformed】成另一个新变量费希尔的 z（记作 z_r）时，该分布就会变得正态，ln 为自然对数【Logarithm】函数。因此，要检验虚无假设 ρ 等于任何非零值（记为 ρ_ϕ），可以运用 z 检验，$z=(z_r-z_\phi)\sqrt{(n-3)}$，其中 z_r 是样本皮尔逊相关系数的转换值，而 z_ϕ 是 ρ_ϕ 的转换值。

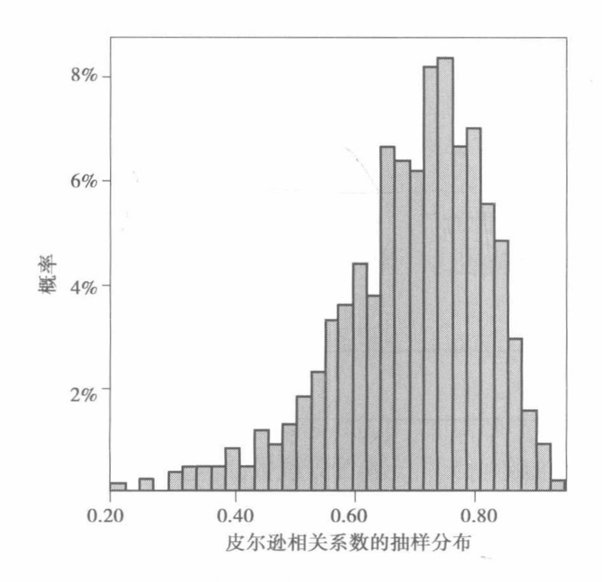


图 3　1 000 个样本的皮尔逊相关系数抽样分布，设 $n=20$ 且 $\rho=0.7$

皮尔逊相关系数的影响因素

皮尔逊相关系数的大小会受到许多因素的影响。首先,只有当两个变量的分布是对称的,并且有相同的形式或形状时,系数的大小才可能是1或-1。它还可能受线性程度和小样本量的影响,尤其是当存在游离值【Outlier】时。此外,由于全距【Range】的限制,系数大小可能会增大或者减小。当任一变量的全距在分布的任一端被限制,皮尔森相关系数往往会减小。相反的,当分布的中矩(middle range)受到限制时,则它会增大。此外,当系数的计算是基于不均匀样本时,它的大小可能是误导性的。最后,它还会受到变量测量误差的影响。测量变量包含越多的误差【Error】,变量之间的相关就越小。

——Peter Y. Chen

Autumn D. Krauss

(叶鹏飞译　高勇校)

参考文献

Chen, P. Y., & Popovich, P. M. (2002). *Correlation: Parametric and nonparametric measures*. Thousand Oaks, CA: Sage.

Hays, W. L. (1994). *Statistics* (5th ed.). New York: Harcourt Brace College.

Stigler, S. M. (1986). *The history of statistics*. Cambridge, MA: Belknap Press of Harvard University Press.

Wherry, R. J., Sr. (1984). *Contributions to correlational analysis*. New York: Academic Press.

皮尔逊 *r* 系数(Pearson's *r*)

参见皮尔逊相关系数【Pearson's Correlation Coeffieient】。

百分比频次分布(Percentage Frequency Distribution)

百分比频次分布是相对频次分布【Relative Frequency Distribution】的一种形式,其中*Y*轴是变量【Variable】每个类别所包含个案的百分比(而不是绝对次数)刻度。例如,在美国,某大学某个春季学期注册的22 000名本科生中,30%是大学一年级新生,27%是大学二年级学生,20%是大学三年级学生,以及23%是最高年级的学生。

——Tim Futing Liao

(叶鹏飞译　高勇校)

*也可参见相对频次分布【Relative Frequency Distribution】。

百分位数(Percentile)

与变量的绝对数值相反,百分位数可以被定义为一种基于百分比的测量系统。百分比建立在100的尺度(即有100个分割出来的部分)基础上,每个部分是一个百分比或百分位数。

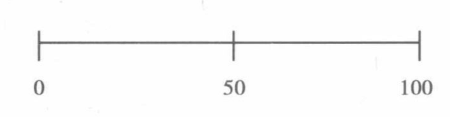


图1　一条线段中的部分

图1中的线段可以分为100个相等的部分。到中间一半是直线相等的50个部分。这个位置是直线总长度的50%,也被称为第50百分位数。如果我们停在第75个部分,这是直线总长度的75%,也就是第75百分位数。如果停在第62个部分,是直线总长度的62%,也就是第62百分位数。

很显然,当我们所测量的变量按照100份进行测量时,应用绝对值测量系统和百分比测量系统得出的结果完全是相同的。也就是说,绝对值测量系统有100个相等的部分,而百分比测量系统也是如此。一次考试得分为42,代表百分比分数为42%,而且可以被认为是第42百分位数。

当两种测量系统不一样时会发生什么?如果试卷的满分为30,我们怎样得出百分位数?答案是将绝对值分数换算成百分比。

例如,应试者1在满分为30分的统计学考试中获得25分。

我们需要把30分中的实际分数25转换成百分比,这时用下面的公式来进行:

百分比分数=(实际分数/满分)×100/1

从上面的例子来计算,结果如下(记住数学运算规则——首先进行括号中的计算):

百分比分数=(25/30)×100/1=83.33%或第83.33百分位数

25的得分是第83百分位数(四舍五入到最近的整数)。应试者和主考官现在就能更好地作出解释和得出结论。

我们可以通过对百分位数进行等级排序来作稍为深入的百分位数分析。要做到这一点,需要知道由我们感兴趣变量的得分引起的频次。一个百分位数的等级是一个分布【Distribution】中得分的百分比,等于或大于一个特定的绝对分值。

如果这个统计学考试的主考官知道参加考试的46人中有40人的得分等于或大于25,那么可以通过将绝对数变换为百分比来计算百分位数的等级,即

(40/46)×100/1=86.95%

也就是说,所有的学生中有86.95%的学生取得了25分或以上的成绩,这个学生是在第87百分位数(四舍五入到最接近的整数)。百分位数的等级为87。

通过把一个变量的分布【Distribution】划分成百分位数,就有可能将每个百分位数作为一组来分析,检验这些组之间的异同。通常,把变量按照十分位【Decile】和四分位【Quartile】分组来进行这类分析。或者,通过计算百分位数的等级【Rank】,我们可以表明在这个研究中,某个人相对【Relative】于其他人的等级。

假设我们已经收集了关于薪酬、性别、商品和服务上的消费等数据。通过将薪酬范围变换为百分位数,可以分析每个百分位数上的组群来探究受访者在性别和薪酬上的异同。此外,还可以研究薪酬与商品和服务消费模式之间的关系。运用十分位和四分位,我们可以减少组数,扩大薪酬间距,这样更容易区分出异同。

排序百分位数让我们可以清晰地看到我们在100位当中所占据的位置,第100百分位数为最高位置。在上述统计学考试的例子中,某学生取得25分(相当于第83百分位数),这告诉我们,该学生在这次考试中排列的等级是83。我们可以产生所有学生的百分位数等级,由此可以看到这名学生与他人相比较的情况。或者,我们可以选择一名学生个体,计算他所有考试的百分位数等级,由此可以看到这名学生的各个成绩等级。这让我们可以发现学生个体的优势与不足。

——Paul E. Pye

(叶鹏飞译　高勇校)

参考文献

Collican, H. (1999). *Research methods and statistics in psychology*. London: Hodder & Staughton.

Saunders, M. N. K., & Cooper, S. A. (1993). *Understanding business statistics: An active learning approach*. London: DP Publications.

时期效应(Period Effects)

个人主体和其他实体包括如城市或国家,可以被认为存在于两个时间维度,即主观的和客观的。一个个体的主观时间维度始于出生,这对于其他实体来讲也是相似的。客观时间维度以个体或实体外部的事件来测量,可以追溯到宇宙的开始,或者更适当地说,是一项研究的开始。我们称主观时间维度为年龄(age),客观时间维度为历史(history)。主观时间开始于同一客观时间的实体属于相同的同期群(cohort)。

在纵贯研究【Longitudinal Research】中,变量的变化模式和变量间的关系模式都是从年龄和历史两个方面来进行研究的。当忽略历史时间维度,年龄对变量或变量间关系的影响被称作发展效应或年龄效应。年龄效应包括语言习得和其他认知能力的发展。当忽略年龄时间维度,生活在某个历史时期对变量或变量间关系的影响被称为时期效应或者历史效应。时期效应包括计算机、手机和其他设备的应用,它们依赖于历史时期中的技术水平,而不是这段时期中的个体年龄。它还依赖于政治和军事事件所塑造出来的态度,诸如在珍珠港被炸或世贸中心被摧毁后,在美国表达更多的爱国态度或更频繁地展示美国国旗。在某个时期处于某个年龄(换句话来说,在某一年出生)的影响被称为同期群效应(cohort effect)。在纵贯研究中,一个首要的问题是分离年龄、时期和同期群的影响(Menard, 2002)。因为如果我们使用年龄、时期和同期群作为预测因子,并假设它们与因变量存在线性关系,那么当同期群是通过出生年份来测量时,根据相关方程:时期=年龄+同期群,这些预测因子就会是完全共线的。

梅纳德(Menard,2002)就时期效应与年龄和同期群效应分离的不同方法进行了总结。霍布克莱夫特,门肯和普特斯顿(Hobcraft, Menken & Preston, 1982)提出:“时期效应背后是这样一种解释方式,即‘是什么年份’并不重要,真正重要的是那一年中或者是那年之前某个重大时期所发生的事件。”梅纳德(Menard,2002)建议,在这个意义上,纵贯分析的时期效应应该被替换为历史事件或它们代表的趋势。在总体水平上,还可通过考察与年龄有关的指标(age-specific indicators)来分离时期效应与年龄或同期群效应。这在社会指标研究中经常被采用,例如,运用婴儿死亡率的趋势或16~64岁的劳动力参与率,作为国民健康和雇用条件趋势的指标。

——Scott Menard

(叶鹏飞译　高勇校)

参考文献

Hobcraft, J., Menken, J., & Preston, S. (1982). Age, period, and cohort effects in demography: A review. *Population Index*, *42*, 4-43.

Menard, S. (2002). *Longitudinal research* (2nd ed.). Thousand Oaks, CA: Sage.

周期性(Periodicity)

周期性变化是一个时间概念,长期以来让普通人深感困惑,给哲学家带来启发,让数学家感到受挫。几千年来,饥荒与繁荣轮替,战争与和平,冰河时代与气候变暖,以及其他,都以一种循环的方式存在。一次循环是一种“周期性和规律性的不断重复”模式(Granger,1989, p. 93)。借助于三角函数,一个时间序列 z_t 的循环分量(the cyclical component)可以表示如下:

$$z_t = A\ \sin(2\pi ft + \varphi) + u_t$$

其中,A 为振幅,f 为频率,φ 为周期性变化的初始相位,u_t 表示白噪声【White-Noise】误差。振幅表示一个循环期间(从峰值到谷底的距离)某个系列摆动的宽度,而频率 f 测量循环运动的速度,即

$$f = \frac{1}{P}$$

其中,P 是一个循环周期,即完成一个循环所需要的时间单位数。因此,一个 4 年循环会以 0.25(每个单位时间里的循环数)的频率运行。对于任何给定的数据,可检测到的最短循环应有两个时间单位的周期。这对应一个 0.5 的频率,是最快的一种可能。任何小于两个时间单位的循环是检测不到的。循环越长,也就是它使用的时间单位越多,则频率越低。没有循环的时间序列,其频率为零,对应着一个无限长的周期。

频谱分析【Spectral Analysis】是识别时间序列中循环分量的主要工具(Gottman, 1981;Jenkins & Watts,1968)。频谱密度函数的典型图表,特征是频谱密度沿纵轴分布,频率(范围从 0~0.5)在横轴上。一个 4 年循环的时间序列(年是该序列的时间单位),会在 1/4 或 0.25 频率处显示出密度峰值。问题是,很少有时间序列会以它们这种频谱密度函数如此完美循环的方式发生。

除了严格的季节性现象,许多明显的循环并不是根据恒定的规则重复自身。例如,经济的衰退,并不是每 7 年发生一次。每一次造成的经济伤害也是不相同的。虽然繁荣与萧条依然交替出现,但这种“循环”周期是不固定的,而且“幅度”也不固定。缺乏恒定性,破坏了如正弦曲线之类的确定模型【Deterministic Model】的效用。作为替代方法,耶尔(Yule, 1927/1971)提出了概率【Probabilisty】模型,用于循环周期和振幅不规则的时间序列。他这一开创性的想法是构建一个二阶自回归模型,模拟 175 年的跨度中,太阳黑子观察值的周期性波动,尽管这个周期不规则。耶尔的 AR(2)模型的形式如下:

$$z_t = b_1 z_{t-1} - b_2 z_{t-2} + u_t$$

这个模型只有两个参数【Parameter】和不超过两个滞后【Lag】,成功地估计了太阳黑子波动的周期(10.6 年)。该模型还提供了对不断影响波动幅度和相位的随机干扰的估计。Box-Jenkins 建模法【Box-Jenkins Modeling】扩展了概率模型的应用,用一组针对季节性波动的特定模型来识别和估计时间序列中的周期分量。

——Helmut Norpoth

(叶鹏飞译 高勇校)

参考文献

Gottman, J. M. (1981). *Time-series analysis: A comprehensive introduction for social scientists*. Cambridge, UK: Cambridge University Press.

Granger, C. W. J. (1989). *Forecasting in business and economics*. Boston: Academic Press.

Jenkins, G. M., & Watts, D. G. (1968). *Spectral analysis and its applications*. San Francisco: Holden-Day.

Yule, G. U. (1971). On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers. In A. Stuart & M. Kendall (Eds.), *Statistical papers of George Udny Yule* (pp. 389-420). New York: Hafner. (Original work published 1927)

排列检验(Permutation Test)

排列检验是一种运用重抽样方法的非参数【Nonparametric】检验。其基本步骤如下:

①设定一个假设。

②选择一个检验统计量并计算原始样本的统计量 $T(T_0)$。

③重新排列样本,再计算重新排列后的样本的检验统计量($T_1, T_2, \cdots, T_x$)。重复该步骤,直到穷尽所有可能的排列方式,或者直到排列数量足够多,以此得到检验统计量的分布。

④将 T_0 与从 T_1 到 T_x 的排列分布进行比较,并计算 p 值。然后判断是否接受假设。

排列检验与自举法【Bootstrapping】相似:同样是基于当前数据进行大量计算机运算。但两种方法的重抽样方法不同:在排列检验中,我们用重排样本来构建分布,但在自举法中,我们从整体样本中有放回地抽样得到一系列样本。

当需要比较来自两个总体的样本时,排列检验的唯一要求是,在虚无假设下,实验组数据的分布应与对照组数据的分布相一致。

排列检验提供一个精确的统计量,而非近似的。当样本量非常大时,排列检验与参数检验一样强效,而在样本量较小的情况下则比参数检验更有效。因此,当难以获得大样本时,比如一个小群体中的社会经济数据,排列检验是非常有用的。

让我们来看一种控制血压的新药的假设数据。新药物的效果与安慰剂进行比较。数据收集自 20 个血压在 165 上下的个体。这 20 人被均匀地分成两组(每组 10 人)。第一组 10 人使用安慰剂,剩余 10 人使用新药。使用一段时间后,安慰剂组的血压数据为{132,173,162,168,140,165,190,199,173,165},实验组的血压数据为{150,124,133,162,190,132,158,144,165,152}。这两组数据能否有证据表明,新药能够真正有效地降低血压?我们计算两组的平均值。对于安慰剂组,平均血压是 166.7;而对于实验组,平均血压是 151.0。差别是 15.7。现在,我们转置这 20 人的数据。共有 $\binom{20}{10}=$ 184 756 种不同的组合。我们不必考虑所有可能情况,因为这样做过于费时。所以,我们随机重排这 20 个数据足够多次,比如说 100 次,并计算转置后数据的前 10 个样本均值和后 10 个样本均值的差。然后,计算原始样本的两组均值之差为 15.7,大于转置样本两组均值之差的次数。如果在 100 次转置中出现了 4 次($p=0.04$),那么在 $\alpha<0.05$的显著性水平下,我们得到的结论是新药能有效地降低血压。

排列检验可以应用于包括各种数据的众多社会科学研究;一个这样的例子是重要的社会经济(网络)数据。我们定义 i 和 j 之间的对称度为 $s_{ij}=|A_{ij}-A_{ji}|$,其中 A 是一个关系的邻接矩阵,A_{ij}是从 i 到 j 的关系程度。

较大的 s 意味着该关系偏离对称性或对称性较小。我们还定义整个群体中的对称性为:

$$S = \frac{1}{\binom{N}{2}} \sum_{i=1}^{N-1} \sum_{j>i}^{N} s_{ij}$$

我们可以计算 S,但由于 S 的分布是未知的,我们不知道与其他组的对称性相比,它是大还是小。因此,我们可以对检验统计量 S 的对称性进行排列检验。

使用排列检验处理社会经济数据时有几个实践方针:首先,我们不转置对角值(diagonal value),因为它们通常有一个理论上的定义值,如 0 或 1。其次,我们需要决定如何转置矩阵。例如,我们可以选择转置整个矩阵,或在每一行中进行转置等,这取决于你的兴趣。一般情况下,通过限制更多的排列,我们可以进行更好的检验。在 S 对称性的例子中,在各行内进行重新排列更有意义,因为我们可以计算 s(由此估计 S),同时保留每个个体的属性,诸如均值和方差。

对于对称性和传递性(transitivity)检验,相关网站可以提供计算机程序 PermNet。

你可以用一些常见的统计软件包执行排列检验。SAS 软件可以使用一个宏程序。SPSS,StatXact 和 LogXac 的精确检验程序包都进行排列检验。

——Ryuhei Tsuji

(叶鹏飞译 高勇校)

参考文献

Boyd, J. P., & Jonas, K. J. (2001). Are social equivalences ever regular? Permutation and exact tests. *Social Networks*, *23*, 87-123.

Good, P. (1994). *Permutation tests: A practical guide to resampling methods for testing hypotheses*. New York: Springer.

Tsuji, R. (1999). Trusting behavior in prisoner's dilemma: The effects of social networks. *Dissertation Abstracts International Section A: Humanities & Social Sciences*, *60*(5-A), 1774. (UMI No. 9929113)

个人文档(Personal Documents)

参见文档类型【Documents,Types of】。

现象学(Phenomenology)

从最有影响的哲学意义上来讲,现象学是关于事物如何出现在意识中的描述性研究,目的通常是识别出决定我们对世界如何体验的基本结构。在社会科学方法论的语境下,这个术语常常意指一种方法,主要关注被研究的人们如何体验世界。它反对使用任何直接的手段来评估这种体验(例如,通过与科学知识相比较得出是真或假)乃至因果性地解释人们为什么用他们正在做的方式来体验世界。从现象学的视角看,这两种趋势都不能把握人们关于自身和世界理解的复杂性和内在逻辑。现象社会学与人类学相对主义有共同之处,坚持"用行动者自己的词语"研究其他文化;以及运用符号互动论【Symbolic Interactionism】的方法,强调人们持续地理解发生的情况,而且这个过

程产生了多样的视角和“世界”。

哲学中的现象学运动是由胡塞尔在 19 世纪末期发起的,后由海德格尔、梅洛·庞蒂、萨特,以及(对社会科学最有影响的)舒茨发展成多个方向。现象学的基本论点是意识的意向性(intentionality):我们总是会意识到某些事。其含义是体验主体与被体验的客体本质上是相关联的;它们彼此不能分离而存在。现象学的影响格外吸引了这样一些社会科学家,他们反对实证主义【Positivism】将自然科学方法作为理性探究的唯一模式,以及所有推理性的社会理论化形式。现象学既是一种针对它们的反对观点,也充当了一种适合于严格的调查研究的替代性模型。

在社会科学中最有影响的一个现象学概念是被称作“自然态度”。胡塞尔认为,人们并非是以传统经验主义者【Empiricists】所宣称的那样对世界的体验是一些零散的感觉。在绝大多数时间里,人们体验着一个熟悉的世界,这个世界包含的各种各样的客体都是可辨认的,人的特征和行为都是已知的(除非另有说明)。虽然我们对世界的体验中还有不确定的领域,而且有些我们的期望得不到满足,但是这些问题背后都有一些我们视为理所当然的背景。现象学探究的任务,特别是由后来的现象学家(如舒茨)所构想的,是描述自然态度的建构特征:我们将这个世界体验为一个理所当然的和共享的现实的社会过程。他们认为,如果我们要为任何一种科学提供可靠基础,这就必须要得到解释。

凭借伯杰和拉克曼(Berger & Luckmann,1967)的工作——他们激发了很多种社会建构论【Constructionism】以及常人方法论【Ethnomethodology】,现象学或许已经对社会研究产生了最直接的影响。

——Martyn Hammersley

(叶鹏飞译　高勇校)

参考文献

Berger, P., & Luckmann, T. (1967). *The social construction of reality*. Harmondsworth, UK: Penguin.

Embree, L., Behnke, E. A., Carr, D., Evans, J. C., Huertas-Jourda, J., Kockelmans, J. J., et al. (Eds.). (1996). *The encyclopedia of phenomenology*. Dordrecht, The Netherlands: Kluwer.

Hammond, M., Howarth, J., & Keat, R. (1991). *Understanding phenomenology*. Oxford, UK: Blackwell.

Schutz, A. (1973). *Collected papers* (4 vols.). Dordrecht, The Netherlands: Kluwer.

Phi 系数(Phi Coefficient)

参见对称量度【Symmetric Measures】。

社会研究哲学(Philosophy of Social Research)

社会研究哲学是社会科学哲学【Philosophy of Social Science】的一个分支。

它不但涉及社会研究的不同部分及其关系，还包含了对社会研究实践的一种批判性检视。

社会研究者在设计和实施社会研究的过程中,往往面临很多选择。这包括研究难题;探究这一难题的研究问题【Research Question】;回答这些问题的研究策略;伴随这些策略的社会调查方法;指导调查的概念和理论;资料的来源、形式和类型;从这些资料来源中进行选择的方法;以及收集和分析资料以回答研究问题的方法(Blaikie, 2000)。在研究过程中的每个节点,要认真评估每个可能的选项,最终有理有据地做出选择,就都会涉及哲学议题。对于社会研究者来说,通过简单地不加批判地遵循与一种特定范式【Paradigm】相关联的做法来回避这些议题已变得越来越困难。

哲学议题

与研究实践中的选择有关的哲学议题,包括各种类型的本体论和认识论假设、解释【Explanation】的本质、理论【Theory】类型和建构方法、语言的角色、研究者的角色、客观性【Objectivity】和真理,以及结果是否具有跨时空的普适性(Blaikie, 2000)。这些议题是相互联系的,根据若干议题作出的选择也会影响根据其他议题作出的选择。这里只对这些议题进行简要讨论,更详尽的分析请参见本书的其他相关词条。

本体论和认识论假设

理论家和研究者们认为这个世界中真实存在的到底是哪些东西,我们又如何认识这些东西,这是科学哲学家要去考察的核心内容。无论是否认识到,社会研究者都不得不提出关于这两个问题的假设,以推进他们的工作。对此有各种各样的立场。为了捍卫已选择的立场,社会研究者需要了解支持和反对每种立场的争论,并且能将自己选择的立场正当化(参见本体论【Ontology】和认识论【Epistemology】关于立场问题的讨论)。

理论类型及其建构

哲学对理论采取的形式和理论产生的过程给予了大量关注。在社会科学中,区分两种类型的理论是非常有用的:理论家理论和研究者理论。前者运用抽象的概念和观点在宏观和微观两个层次上理解社会生活。后者在范围上更具体,是在社会研究的情境中发展和/或检验的。尽管这两类理论是相互关联的,但它们的形式和抽象层次是不同的(Blaikie, 2000, pp. 142-143, 159-163)。这里涉及的主要是研究者理论。

有些研究者理论是由相互关联成网络的一般规律所组成,这些规律是从资料中归纳【Induction】而来。其他研究者理论则表现为一组具有不同普遍性层次的命题形式,这些命题来自演绎的论点。它们的起源无关紧要,但它们的经验检验被看成是至关重要的(参见证伪主义【Falsificationism】)。有些理论是由关于生成结构或机制的描述所组成的,被认为可以解释一些现象,并且是通过应用对问题的合乎情理的想象(informed imagination)而获得的(参见回溯推理【Retroduction】)。其他由丰富的描述或抽象组成的理论,可能的形式包括行动者的理想类型、行动步骤和情境。它们都来源于社会行动者的概念和阐释(参见常情诠释【Abduction】和阐释论【Interpretivism】)。社会科学哲学家已经对所有的四种形式理论建构都提出了论证和批评。

解释的本质

社会科学中最难解决的问题之一是如何解释社会生活。争论的中心集中在解释的来源上，是基于社会结构还是基于个体动机，即整体主义还是个体主义。问题就是这个来源是从社会行动者的外部还是他们的内部去寻找。

相关的争论涉及解释的类型，需要区分因果解释和动机解释。前者可以采取多种形式，遵循自然科学中使用的解释逻辑，但后者不认为自然科学的逻辑适合社会科学，而代之以社会行动者赋予行动的道理或动机。道理或动机，是否可以作为解释，还是仅仅提供理解，是某些争论的焦点。

无论是在自然科学还是社会科学中，因果解释都具有三种主要形式：模式解释、演绎解释和回溯推理解释。在模式解释（参见归纳【Induction】和实证论【Positivism】）中，一个确定的普遍化的社会生活规律可以提供最根本的解释。例如，如果确定了“少年罪犯出自破裂家庭”，就可以用它来解释为什么一些年轻人会犯罪，原因是他们的家庭背景。演绎解释（参见证伪主义【Falsificationism】）从一个或多个抽象的理论假设出发，并运用演绎逻辑和其他命题，得出一个推论，将它作为一个假设或作为对一个研究问题“为什么”的回答。回溯推理解释（参见回溯推理【Retroduction】和批判实在论【Critical Realism】）用背后的因果结构或机制来解释我们所观察的现象的特征。

语言的角色

这里的哲学议题是科学语言与世界上的事物之间如果有关系，应该是一种什么样的关系。在这个问题上采取的立场依赖于采用的本体论假设。例如，在实证论【Positivism】中，运用理论中立（theory-neutral）的观察语言被认为是不成问题的，但是在证伪主义【Falsificationism】中，这是被拒绝的，原因是所有观察都依赖于理论。

在实证论中，语言被看成是描述世界的媒介，并且假定语言的结构形式与语言用来描述的对象世界之间具有同构性。然而，在阐释论【Interpretivism】的某些分支中，语言被认为是日常社会行动实践的媒介。这里的议题是这种非专业性的语言与社会科学语言之间的关系，一些作者主张，后者应当通过常情诠释【Abduction】从前者中产生。选择在于是对社会世界强加技术语言还是从日常语言中引申出技术语言。

研究者的角色

社会研究者如何对待研究过程和参与者，对此有各种不同基本立场。有人主张彻底超脱，有人主张全心投入。社会科学哲学家已对这些立场的可行性及其对于研究结果的影响进行了批判性的检视。

每一种立场都与社会调查和研究策略的特定方法相关。独立观察者（detached observer）的角色关注客观性【Objectivity】，并与实证论以及归纳和演绎逻辑有关。忠实报告者（faithful reporter）的角色关注社会行动者的观点，并与阐释论的某些分支有关。语言媒介者（mediator of languages）的角色关注理解的产生，并与诠释学【Hermeneutics】的某些分支以及常情诠释的逻辑有关。反思性伙伴（reflective partner）的角色关注解放，并与批判理论和女性主义有关。最后，对话性促进者（dialogic facilitator）的角色与后现代主义

【Postmodernism】有关,并关注通过允许被压抑的各种“声音”来减少研究者的权威影响力。

客观性

没有哪一个与社会研究相关的概念比客观性更能引发人们的情感和困惑。研究者的目的通常是希望得到客观的结果,而当他们得不到时会经常受到批评。一些研究方法因为看上去比较客观而受到赞赏,而其他一些方法则因为不客观而受到批评。这里的哲学议题是:“客观”到底意味着什么?什么样的方法和结果才是客观的?

在实证论中,客观意味着坚持事实与价值的分离,也就是说,不允许研究的价值和偏见影响数据。这是一个具有欺骗性的简单想法,而且被证明是不可能实现的。首先,“事实”的概念非常复杂,并且事实与价值之间的区别并不是界限分明的。对某些哲学家来说,坚持事实与价值的分离,可能意味着在研究的议题上不偏袒任何一方。但是,一些研究者喜欢某种特定的观点(在女性主义中,女性的观点)。价值本身也可以是一个研究的主题。

对一些哲学家而言,客观性可以通过运用那些被科学共同体认为是适合的方法来获得。当然,这本身就意味着一种特定的价值标准。也有人把客观性等同于结果是可重复的。其他研究者能够重复研究的程序,从而实现对结果的验证。然而,特定方法的可接受性是评判的标准。

一些阐释的社会科学家回避对客观性的关注,认为研究者应当尽可能是主观的。意思是研究者应将自身完全融入他们所研究的人群的生活中,因为这似乎是理解另一种生活形式的唯一方式。他们的目标是理解那一生活形式的真实意义,即一种忠实于社会行动者观点的意义。

真理

哲学家发展出若干理论来捍卫他们关于“真理”本质的观点。真理符合论(the correspondence theory of truth)要求从“客观的”观察中获得证据。它假定一位没有偏见的观察者将能够看到世界的本来面目,并且能够用理论中立的语言报告这些观察结果。它假定一个理论为真的基础是该理论与事实一致(参见实证论【Positivism】)。当然,要实现这一点,必须就什么是事实达成一致。真理暂时论(the tentative theory of truth)认为,观察的理论依赖和证伪主义逻辑本质意味着从理论中演绎而来的检验永远不会获知真理。尽管发现真理是目的所在,但是演绎理论检验的结果通常要服从于未来的修正。与此相反,真理共识论(the consensus theory of truth)是建立在理性探讨的理念基础上,前提是要摆脱了所有限制和诱导的影响。尽管证据可能是重要的,但是理性的争论可以确认真理。然后是真理实用论(the pragmatic theory of truth),认为如果某个事物是有用的,能够融合到一个群体的经验中并能得到验证,那么它就是真的。在符合论中,真理被认为是永久性建立起来的。在共识论中,任何人无偏的合乎逻辑的论证都会获得相同的真理。然而,在暂时论中,真理被认为是随着检验过程而改变的,在实用论中则是随着经验的改变而变化。最后,在真理相对论(the relativist theory of truth)中,存在众多的真理,因为它假定存在多样的社会现实。正如没有人拥有特权的或中立的立场来判定多样的真理,我们不可能决定哪一个是“最真实的”。这种立场范

围使真理的概念更少稳定性而不是像常识认为的那样。

概推

所有科学工作者似乎都有一种愿望,希望超越他们得出研究结果的情境来推广他们的成果,这样他们的理论应用就能跨越空间和时间(参见概推性【Generalization】)。这对实证论而言是特别适当的,其中使用的是归纳逻辑。尽管事实是建立在有限的观察基础上,但目标却是要建构普遍规律,比如"少年罪犯出自破裂家庭"。很多哲学家认为这一目标是不可能实现的,并建议少一点这种过于宏大的愿望。例如,证伪主义者承认其理论的暂时性,认为理论可能会在将来被更好的理论替代。此外,在他们的理论中,不仅会以一般形式来陈述一些命题,而且也包括了影响理论适用性的限制条件。事实上,很多科学活动正是致力于探索这类条件。例如,尽管"少年罪犯出自破裂家庭"这个一般命题可能列在一个演绎理论的前面,但诸如"在某些破裂家庭中,儿童没有充分社会化到社会规范中"这类限制条件,将这一理论限制在特定类型的破裂家庭。批判实在论【Critical Realism】对潜在机制的探寻,是建立在这一假设的基础上,即它们是现实的而不只是科学家的创造。然而,这些研究者认为他们的任务不仅是要发现适当的机制,而且是要确立这些机制运作的条件和情境。机制可以是普遍的,但它们的运作过程通常依赖于环境。因此,这些理论通常受限于时间和空间。在阐释论中,普遍接受的是,任何关于社会生活的社会科学描述必须依据时间和空间加以限定。任何理解必须限制在其本身常情诠释而出的那个情境。同一类型的理解在其他社会情境中是否也是适当的,这将依赖于更进一步的调查研究。无论如何,所有理解都是有时效性的。

概推导致相当大困惑的一个方面,是人们经常混淆理论命题的概推与抽样调查中的概推(参见抽样调查【Survey】)。如果一个概率样本【Probability Sample】来自一个确定的总体【Population】,并且回答率是符合要求的,那么从样本中等到的结果就能够在统计上推广到从中抽取样本的总体中。在这个总体之外,所有推广性必须依据其他的证据和论据基础,而不是统计程序。就此而言,超出抽样总体的概化,本质上等同于从非概率样本【Nonprobability Sample】和个案研究【Case Study】中进行概括。

社会研究的几个方面

社会研究实践的若干方面与前面讨论的几个议题有关。这里讨论的是如何从各种选项中进行选择。

社会探究的途径

在抽象层次上,社会研究有着不同的理论和方法论视角,也就是社会探究的不同路数。这些路数可以用不同方式来表述,通常包括实证论、证伪主义(或批判理性主义)、诠释学、阐释论、批判理论、批判实在论、结构化理论和女性主义(Blaikie, 1993)。每一种途径都有自己的本体论和认识论假定(参见 Blaikie, 1993, pp. 94-101)。选择哪种路数和研究策略,会对研究设计【Research Design】计划和具体研究项目的实施产生较大的影响。

研究策略

社会研究中可以采取四种基本策略,它

们是基于归纳、演绎、回溯推理和常情诠释的逻辑(Blaikie, 1993, 2000)。每一种研究策略都有一个不同的出发点,并得出不同的结果。归纳策略是从数据收集出发,从中作出推广性结论,用此作为基本解释。演绎策略是从对研究问题的"为什么"提供一种可能答案的某个理论出发。从该理论逻辑演绎而出的结论,是我们需要的答案。该理论要在研究问题的情境中通过收集相关数据来进行检验。回溯推理策略是从一个假定的机制模型出发,该模型能解释调查研究中现象的发生。然后必须进行研究以论证这一机制的存在。常情诠释策略是从非专业性的概念和意义出发,这些概念和意义包含在社会行动者关于与研究问题相关的行动的陈情【Account】中。然后这些陈述会在社会科学家的叙述中运用技术语言进行重新描述。后者提供了一种对现象的理解,并可作为更复杂的理论化的基础。

四种研究策略都有不同的本体论假设。归纳和演绎策略共享相同的本体论假设(参见实证主义【Positivism】、归纳【Induction】、证伪主义【Falsificationism】和演绎【Deduction】);回溯推理策略使用的现实概念是分层次的,明确地区分了现实(domains of the actual)(我们能观察到的东西)与实在(the real)(深层结构和机制)(参见批判实在论【Critical Realism】和回溯推理【Retroduction】)。常情诠释策略采取了一种关于社会现实的社会建构论观点(参见阐释论【Interpretivism】和常情诠释【Abduction】)。上述差异会影响到研究中的其他选择和结果阐释,因此了解不同策略之间的差异非常重要。

归纳策略对于回答"是什么"的研究问题非常有用,但基于资料推广已被证明不能充分地回答"为什么"的问题。演绎和回溯推理策略主要用于回答"为什么"的问题,尽管两者都要求首先建立起对观察特征的描述。常情诠释策略能同时用于回答"是什么"和"为什么"的问题。

对于每一种策略,都有大量哲学上的批评以及对它们优缺点的争论。没有哪一种是十全十美的。最后,社会研究者必须评估这些争论,然后判断哪一种适合手中研究的问题,同时认清它们的局限性。

结论

所有这些表明,对于社会研究中的哲学议题的理解,对于任何社会科学家而言都是必要的。在进行社会研究时必须作出诸多选择,因此社会科学家必须在哲学议题上选择立场。如果对作出这些选择所涉及的哲学议题缺少理解,就会导致幼稚的和自相矛盾的研究实践。

——Norman Blaikie

(叶鹏飞译 高勇校)

参考文献

Blaikie, N. (1993). *Approaches to social enquiry*. Cambridge, UK: Polity.

Blaikie, N. (2000). *Designing social research: The logic of anticipation*. Cambridge, UK: Polity.

Hughes, J. A. (1990). *The philosophy of social research* (2nd ed.). Harlow, UK: Longman.

Williams, M., & May, T. (1996). *Introduction to the philosophy of social research*. London: UCL Press.

社会科学哲学(Philosophy of Social Science)

社会科学哲学是科学哲学的一个分支,是哲学的一个重要领域。哲学学科是很难进行精确定义的。原始的希腊语含义是爱智慧,而今天则与批判性的分析方法相关联。虽然哲学没有自己的具体研究对象,也不生产关于世界的新事实,但哲学家们能够提供对已知事实的新理解。通过揭示概念和理论所指以及探索它们的内涵,他们澄清了知识的首要原则。通过分析我们用来讨论世界的语言,哲学家可以帮助我们理解它。特别是哲学家们对我们认为理所当然的对现实的解读方式提供了一种批判性的反思。

在科学哲学中,关键的分析聚焦于这类主题,即本体论【Ontology】和认识论【Epistemology】的议题"什么是现实"和"什么构成了关于现实的知识"。更明确地说,哲学的这一领域探讨的是科学的本质、用来发展知识的科学探究的逻辑,以及观点背后的理据。

社会科学哲学主要关心三个议题:各种各样的本体论和认识论假设以及这些假设的正当理据何在,这些假设是社会研究者在他们的理论和社会探究方法中采用的;用来产生研究问题【Research Question】答案的各种探究逻辑;某一特定探究逻辑中使用的资料收集和分析方法的适当性。

自然科学哲学中处理的很多议题在社会科学哲学中也非常重要。然而,近100多年,哲学家和方法论学者一直在争论的一个问题是,成功运用于自然科学的方法和探究逻辑,是否也应当应用于社会科学中。该争论的麻烦之处在于,自然科学内部对于何种探究逻辑最有道理也是众说纷纭——归纳【Induction】、演绎【Deduction】还是其他方式?

对所有社会科学家来说,一个基本的选择是,使用这些探究逻辑中的一种,还是使用被一些哲学家认为是更适合社会科学中特定主题的逻辑。我们采用的本体论假设,将决定社会现实是被看成与自然现实在本质上是一致的,还是与自然现实根本不同。如果两者根本不同,我们就得采用不同的逻辑(常情诠释【Abduction】),资料收集和分析方法也会有根本不同。

这一争论的焦点是如何解释人类行为。采用自然科学中运用的逻辑,需要注重建立起因果解释,辨识产生特定情境下这些行为的原因和机制(参见因果关系【Causality】)。然而,那些赞成不同逻辑的社会科学家则认为应当把行动者自己对行为提供的理据作为理解的基础。

因此,社会科学家选择哪些研究问题,选择哪种探究逻辑来回答这些问题,都表明了他们在深刻而复杂的哲学议题上选择的立场。

——Norman Blaikie

(叶鹏飞译　高勇校)

参考文献

Blaikie, N. (1993). *Approaches to social enquiry*. Cambridge, UK: Polity.

Chalmers, A. (1982). *What is this thing called science?* St. Lucia, Australia: University of Queensland Press.

O'Hear, A. (1989). *Introduction to the philosophy of*

science. Oxford, UK: Clarendon.

Rosenberg, A. (1988). *Philosophy of social science*. Oxford, UK: Clarendon.

研究中的相片(Photographs in Research)

相片在社会科学研究中扮演着有限但重要的角色。近年来,这个角色得到了加强。首先,摄影克服了社会科学通常依赖于加总数据分析的倾向,它明显聚焦于具体情况。其次,文化研究和符号理论家已经成为可视化社会层面最具活力的焦点,他们研究如传媒、建筑学和其他可视化社会形式中的支配、抵制和其他文化主题。这不是指运用相片来收集资料,而是用它来分析可视的结构化世界。

相片在社会学中的首次使用,是用来对物质文化或社会互动的不同方面进行展示。在《美国社会学期刊》(*American Journal of Sociology*)的前20卷中,即1896—1916年,社会学家使用相片研究诸如办公室和工厂中的工作条件以及工业化社会中的生活环境等问题。当社会学寻求一种更科学的形式时,相片在半个世纪的时间中都在很大程度上被抛弃。

人类学家也在摄影发展的数十年中大量运用它来记述和分类人类的群体。当人类学抛弃以种族类型的分类为基础的范式时,摄影的重要性就下降了。

摄影在民族志中的全面发展归功于格雷戈里·贝特森和玛格丽特·米德(Bateson & Mead,1942)关于巴厘岛文化的研究。他们通过大量的影像资料研究巴厘岛的社会仪式、日常行为和物质文化。后来,摄影在人类学民族志【Ethnography】中被广泛应用。

霍华德·贝克尔是第一位将摄影与社会学联系起来的现代社会学家(Becker, 1974)。贝克尔注意到,摄影和社会学自其起始阶段,都涉及对社会的批判性观察。20世纪早期的摄影师们,比如海因和雅各布·里斯,记录了城市化和工业化、劳工剥削和移民状况等社会学家所关心的所有主题。贝克尔建议,视觉社会学(visual sociology)应当使得纪录片实践和社会学理论联姻,而相片在社会学中的运用也在关注效度【Validity】、信度【Reliability】和抽样【Sampling】的背景下得到发展。

在此期间,社会学家证明了相片对于研究社会变迁(Rieger, 1996)和其他一些传统社会学主题的实用性。然而,相片仍然难以进入经验研究的主流,而是一种例外。也有人使用相片在宏观或结构层次上来探索社会,比如在德博拉·巴恩(Barndt,1997)关于全球化与西红柿生产和消费的相片拼图/田野研究中。

研究者可以通过相片引导出文化信息。这个过程被称为照片启迪(photo elicitation)法,利用相片在访谈中激发出与文化相关的反应(Harper, 2002)。这些相片可以是研究者在田野观察中制作的,或者是来自历史或个人的档案中。在哈普尔(Harper,2001)近期的研究,其照片来自一个档案馆,是50年前拍出来的,描绘了当时工作、社区生活和家庭互动的日常状态,它们被用来鼓励老年农民去反思农业社区变革的意义。

照片启迪法可以揭示出研究者和被访者头脑中视为理所当然的一些假定。此外,由于照片诱导鼓励基于视觉材料分析的反思,所以记忆和反思的质量要高于单独用语言访谈。最后,照片启迪法挑战了研究者的权威,并且将研究项目重新定义为一种对意

义的合作式探寻,而不是对信息的挖掘。这使照片启迪法明确成为“新民族志”或“后现代研究方法”的新兴实践。

摄影用于社会科学研究,这至少存在三个伦理挑战。首先涉及匿名问题,其次涉及知情同意【Informed Consest】问题,最后是“讲述视觉事实(telling the visual truth)”的压力。

在传统研究中,个体是用文字或者表格和图形来表示的,真实身份被隐去了。因此,研究对象的身份很容易得到保护。然而,在摄影研究中,保护对象的隐私【Privacy】通常是不可能的。因此,摄影研究挑战了社会科学研究的一个基本原则。对这一问题的探讨仍在继续。

视觉研究者【Visual Researchers】面临的另一个问题是知情同意。如果可能,必须让研究对象了解在社会科学研究中公布他们的肖像所可能带来的影响,且他们同意参加的意见应当非常明确并以书面形式确认。然而,当社会科学的摄影者在诸如公园、城市景观和其他非正式互动场所的公共集会拍摄照片时,知情同意就变得不太可能。对这个问题的探讨也还在继续。

视觉研究者必须维护“讲述视觉事实”这一坚定承诺,不能修改图像来改变它们的意义,要呈现一个与对当时的情境进行科学理解的现实相一致的视觉图像现实。这个指导原则是高于一切的伦理规范,超越上面所列的所有技术性注意事项。

新的方法,比如计算机为基础的多媒体、万维网、CD 出版物和虚拟现实,所有这些都为相片作为社会科学方法的创新应用提供了可能。社会研究中相片的未来,将部分依赖于这些新的技术如何能够更好地整合进社会科学方法和思维中。然而,致力于视觉研究的新期刊以及运用相片作为资料和交流的主要形式的书籍不断增加,意味着在社会科学中即将迎来相片大量使用的前景。

——Douglas Harper

(叶鹏飞译　高勇校)

参考文献

Barndt, D. (1997). Zooming out/zooming in: Visualizing globalization. *Visual Sociology*, *12*(2), 5-32.

Bateson, G., & Mead, M. (1942). *Balinese character: A photographic analysis*. New York: New York Graphic Society.

Becker, H. S. (1974). Photography and sociology. *Studies in the Anthropology of Visual Communication*, *1*(1), 3-26.

Harper, D. (2001). *Changing works: Visions of a lost agriculture*. Chicago: University of Chicago Press.

Harper, D. (2002). Talking about pictures: A case for photo elicitation. *Visual Studies*, *17*(1), 13-26.

Rieger, J. (1996). Photographing social change. *Visual Sociology*, *11*(1), 5-49.

饼图(Pie Chart)

饼图是一个圆形被分成多个部分,每个部分表示在总体中的份额或百分比;每一部分被称为这个“饼”的“切片”。条形图可以应用于相同目的,但假如只有相对较少的类别,饼图则更容易让人理解个案是如何在类别间“分配”的(直方图【Histogram】或条形图在需要显示变量的变化趋势时优先于饼图,例如随着时间的变化或比较几个区域)。

例如,图 1 描述了住在一个市镇中心的人们如何评价他们的区域以及什么东西比较好。

饼图还可以用于类别之间的简单对比。这里的条形图也可以应用于相同目的,并且在存在数个不同群组需要比较时更为合适,但饼图能够给出两个群组间的形象的对比,更容易理解并且更直观。例如,图 2 中的饼图,展示了不同族群利用市民咨询局(Citizens' Advice Bureau)的不同原因。市民咨询局是遍布英国的志愿者机构,主要是为那些没有能力的穷人提供建议和引导,如何处理与中央和地方政府、税款、房东、购物相关的以及其他财务和法律事项,并且他们也提供关于债务和财务管理方面的咨询。如图 2 所示,在米德尔斯堡,不出所料,移民和国籍是亚洲人经常提出的问题,但很少被白人关心。白人对住房和福利(政府承担)的关注比亚洲人稍多一点,并且更多的是讨论债务问题。

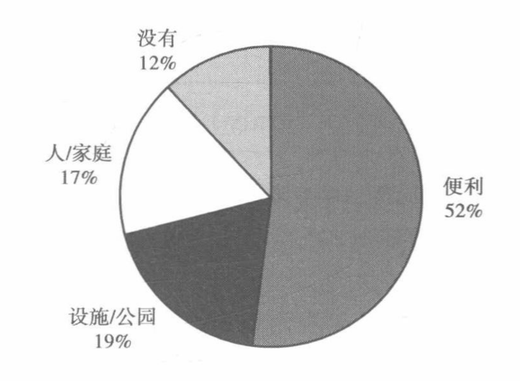


图 1　你的区域在生活方面哪些东西比较好?
资料来源:Hobson-West and Sapsford(2001).

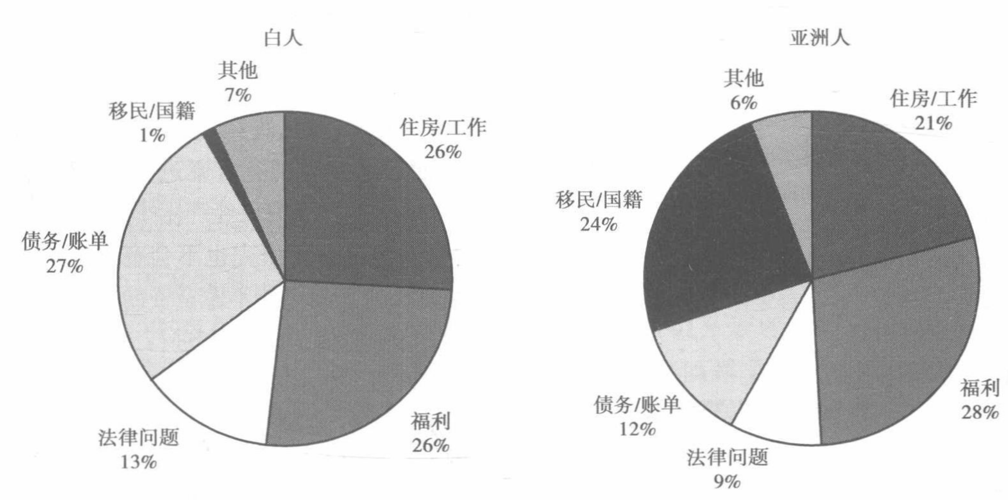


图 2　米德尔斯堡造访市民咨询局的目的
资料来源:McGuinness and Sapsford (2002).

——Roger Sapsford
(叶鹏飞译　高勇校)

参考文献

Hobson-West, P., & Sapsford, R. (2001). *The Middlesbrough Town Centre Study: Final report.* Middlesbrough, UK: University of Teesside, School of Social Sciences.

McGuinness, M., & Sapsford, R. (2002). *Middlesbrough Citizens' Advice Bureau: Report for 2001.* Middlesbrough, UK: University of Teesside, School of Social Sciences.

分段回归(Piecewise Regression)

参见样条回归【Spline Regression】。

预研究(Pilot Study)

预研究是指在主研究的准备阶段实施的可行性研究或小规模研究,以及一个特定研究工具的前测。

良好的预研究可以提升主研究成功的可能性。预研究可以提醒项目失败的可能、与设定方案的偏差或者拟用方法和工具的问题,以及希望发现可能影响研究的地方政治或问题(参见表 1)。

表 1　实施预研究的原因

• 开发和测试研究工具的适当性
• 评估(全面)研究的可行性
• 设计研究方案
• 评估研究方案是否现实和可行
• 建立有效的抽样框和抽样技术
• 评估拟用招募方法的成功可能性
• 识别使用拟用方法可能出现的逻辑问题
• 估计结果的变异性以确定样本量
• 收集原始资料
• 确定研究所需要的资源
• 评估拟用的资料分析技术以发现潜在的问题
• 提出研究问题和/或研究计划
• 对研究者进行研究过程各要素的培训
• 使资助者相信研究团队是称职的和知识渊博的
• 使资助者相信研究是可行的和值得资助的
• 使项目小组相信研究是值得投入的

预研究可以建立在定量(定量研究【Quantitative Research】)和/或定性(定性研究【Qualitative Research】)方法的基础上,而大规模研究可能同时采取多个预研究。一项问卷【Questionnaire】调查的预研究可以从访谈或焦点小组开始,以便设计在问卷中需要询问的议题。然后,再就用词、问题的排序和可能的选项进行试调查。最后,研究过程可能要进行检验,例如分发和回收问卷的不同方式,以及设计用来克服诸如难以记录和回答率(回答偏倚【Response Bias】)等问题的预防程序或安全系统。

实施一项预研究可能会出现的困难:

• 样本污染:曾经参加过预研究的受访者和从未参加过预研究的受访者,回答是有差异的。

• 已经在预研究上进行了大量投入后,研究者很难接受研究有可能是根本不可行的,而只愿意对现有方案进行一定改进。

• 如果预研究已经产生了实质成果,资助者会认为再深入研究也不会有更具原创性结果,那么资助者可能不愿意资助这个主研究。

很少有人对预研究进行完整的报告,只是用来说明使用的研究方法和/或工具的有效性。然而,预研究的过程和结果可能对于其他人使用类似方法或工具开展项目非常有用。研究者有这样的伦理义务,即通过报告所有研究的部分,包括预研究中出现的问题,以充分利用他们的研究经验。

——Edwin R. van Teijlingen

Vanora Hundley

(叶鹏飞译　高勇校)

参考文献

De Vaus, D. A. (1995). *Surveys in social research* (4th ed.). Sydney, Australia: Allen & Unwin.

Peat, J., Mellis, C., Williams, K., & Xuan, W. (2002). *Health science research: A handbook of quantitative methods*. London: Sage.

van Teijlingen, E., & Hundley, V. (2002). The importance of pilot studies. *Nursing Standard*, *16* (40), 33-36.

安慰剂(Placebo)

安慰剂的使用与实验【Experiment】相关。它的主要作用是在医学研究中,引导控制组【Control Group】相信他们得到了治疗,例如一种药物,但事实上治疗是虚假的。安慰剂有时也被使用在不需要产生实际效果的控制组中。由此而言,它的主要角色是作为治疗组或实验组的一个比较点。如果人们因为相信安慰剂确实会有一定的疗效从而导致自身行为或健康情况发生改变,比如安慰剂使人感觉更平静,这种情况被称为"安慰剂效应"。

——Alan Bryman

(叶鹏飞译　高勇校)

计划比较(Planned Comparisons)

参见多重比较【Multiple Comparisons】。

诗学(Poetics)

诗学认为语言既是一种达到各种交流目的的工具,同时也具有自身独有的本质。以语言为中心的行为均具有这种特点,只有根据当时的目的或者被强调,或者被忽略。在诗学中,作品的聚焦点和语言的特殊陈列形式是理解它的关键,而不仅仅是作品的形式(例如,诗化散文,善于用隐喻和寓言;诗可以作为散文或韵文;吟诵,歌唱)或者传递的各种信息(例如,美的、说教的、神话的、仪式化的)。社会科学中对它的兴趣目前越来越浓,尤其是在社会科学将文本作为文化产品的关注中。部分根源于后现代主义【Postmodernism】的争论,部分根源于对文化尽管多元仍存共性的认知不断增加,这种"文学转向(literary turn)"拓展了对符号行为、话语中的意义生产、文本建构、行动中权威的研究,以及更普遍地,与任何形式的经验表达和交流相关的哲学和批判性的问题中诗学领域的研究。它包括广泛的研究主题:批判性对话中观点的相互建构、世界上诗歌的历史和形式(从部落社会到现代剧场)、仪式和世界观以及它们与语言和文化的关系研究,以及自传式民族志和民族志田野工作中自觉收集的档案。它对社会科学具有丰富的含义。

人类学诗学

人类学诗学面临的特殊挑战包括发展既敏感又如实的跨文化翻译技巧,以及一种能够系统地体现启发性的诗意又不牺牲民

族志责任性本质的叙述体裁。我们需要从行动者的视角出发来了解这种本质及其在不同语言和文化中的意义,同时,也需要找到联合和沟通这些差异的途径。假设所有人都对探索和互相移情的沟通持有足够类似的开放态度,人类学诗学试图通过让读者(或听众)接触文本来唤起类似的体验。民族志研究者对这类论点感兴趣,因为他们有助于界定关于自我意识研究和写作的限制条件。这种日益增长的“文学性”,也导致一些社会科学家痴迷于线、标点和空格,简言之,即将诗歌作为另一种表达形式。

诗是迄今为止诗意沟通中最著名的类别。它在概念和内容上否认身心分离,它与人的体验本身最为接近。它同时发生在思想和观点形成的意识层面和内心深处,诗人是有文化的、敏感的、有历练的人。它是一种自我展示、自我建构的探索方式,大体上就像写作和基本的人类活动“讲故事”。它从外来事物、日常事物和我们的梦想中获得动力和素材,在这个过程中创造并占据感觉和意义空间。人类所做的一切都有其意义,从挖红薯到发射火箭到月球,每个意义最终都锚定在感觉上。不过,通常只有诗人才会始终如一地从感觉的角度来书写经历——集中、解码、重构以及准确地讲述,如同“身体力行(embodied)”的参与观察者,富于触觉、嗅觉、味觉、听觉和视觉感,直面日常生活中的烦恼与欢乐、汗水与眼泪、情爱与焦虑。诗人展示出来的关于世界的看法,有时近似一棵植物之上的昆虫之眼,有时又深远到极点和宇宙。在民族志工作中,诗人希望通过相互创作,来揭示这个世界对自己和对其他人的意义,特别是获得无法产生在单一语言或文化内获得的内容。这些内容牵涉不同的模式和困扰,却可以被共同参与者所共享。但是这样做的最终方式,是将通过细致审视将看似陌生者的他者性消解掉。诗强调我们所有人的特质。但是将其纳入社会科学方法的体系则问题重重。它要求这些学科要面对整个复杂的美学领域以及在民族志陈述中富有智慧但却违反直觉的论点,包括语言(和作者的感觉)在科学著述中的地位。

科学与诗学

科学探索让我们通过小规模和大规模的模式能或多或少窥见事物和行为之间的因果关系。它为我们展示的并不是日常现实,不是我们直观体验到的现实。诗人并非仅仅通过抽象概念来书写或谈论关于他们的经历,就像一个人在应用创造出来的科学理论时可能会做的那样,尽可能地聚焦于科学表达的对象而不让语言喧宾夺主;诗人的描述是具体的,他把自己置身于与其他事物的关系框架当中,尤其是物理环境、情感处境与社会环境当中。他们的目标是在不同人的内心世界中寻找到共通的表达方式,采取的方式是利用他们的语言、激发我们内心的事物、发现日常生活中的惊奇,以及暂时将我们带出自身,以便能向我们展示自身的本质,而不是就事论事。诗学使人们可以避免不经思考就接受那些有关文化的标准答案,同时又使人们在这一过程中感受到人性的共通之处。它兼具辩证的视角和比较的视角。

隐喻作为一种探索工具有与此类似之处。祖父是一棵橡树或原子内粒子最好概念化为运动中的台球,这样的思路可能被证明是解开最困扰我们的事物的源头,并且这种洞察力无法通过任何其他方式加以获得。知道这一点,并且有意识地利用类比和隐喻,这将所有的经验诠释黏合起来(对于科学家和诗人都一样,只是科学家对此不承认

而已），诗人认为最富成效的探索形式取决于要解决的问题本质。正如《芬尼根的守灵夜》（*Finnegan's Wake*）不能简化为共济会或河流之旅的主题，不能简化为其形式特征和灵感的关键概要，加利福尼亚的原住民，其跨文化特质分布的统计表达式也无法产生诗性表述或精彩隐喻中所常见的精准表达和修辞效果。这其中的含义非常重要，不仅可以帮助我们解释一个共享世界的文化建构，这个共享世界根据不同的标准和环境既有区分又有联合，而且正如维特根斯坦（Wittgenstein，1974）所言，因为改变了我们的描述语言，从而改变了"分析游戏"本身，包括改变了研究的切入点。因此，诗学并不能被认为是现有定量分析方法的一个简单补充。它是一种完全不同的诠释形式，因此也是认知和叙述的不同形式，它明显属于定性的范围，它绝对不仅仅是其他类型研究的补充。

诗学探究可以为我们提供关于世界的本质以及我们在世界中所处位置的令人振奋和启发性的知识，其中有些是根本无法通过其他手段得到同样的深度和同样的形式。除了确保在自我意识基础上进行诠释，它的目的还在于避免思考的武断结束，鼓励研究和报告中的创造性。需要提醒的是，尽管存在许多共同特征，但科学和诗学所做的工作不同；没有任何单一的流派或方法可以捕获它的全貌；我们所说的一切都不可能被套进全新的事物中；而且在某种程度上，对我们所有人来说，经验和表达的领域都是受文化所限的和不完整的。如果我们总是要理解我们所研究人群的日常现实——他们所知道和解释的宇宙以及作为有情众生的行动，那么缓和或解决此类问题（例如，通过超越分析类别的方法，这些类别中只存在于研究者自身的思想和认同中）就很重要。忽略这些现实，我们就无法理解自身，我们是如何社会性地和符号性地被表达出来，我们是如何在我们的意识中和与他人的比较中将我们的自我建构成富于意义的实体，以及如何在同一个星球上频繁的互动过程中促进负责任的行动。认为某种方法在任何时候都优于其他方法，是真理的唯一来源，这就相当于混淆了苹果和榔头——认为存在适合所有工作的一种工具。包容是更好的。民族学科学和人类学诗学的结合，能够更可靠而完整地解释我们这样一种存在：一种创造意义也渴求意义的生物。

——Ivan Brady

（叶鹏飞译　高勇校）

参考文献

Brady, I. (2000). Anthropological poetics. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 949-979). Thousand Oaks, CA: Sage.

Brady, I. (2003). *The time at Darwin's Reef: Poetic explorations in anthropology and history*. Walnut Creek, CA: AltaMira.

Rothenberg, J., & Rothenberg, D. (Eds.). (1983). *Symposium of the whole: A range of discourse toward an ethnopoetics*. Berkeley: University of California Press.

Tedlock, D. (1999). Poetry and ethnography: A dialogical approach. *Anthropology and Humanism*, *24*(2), 155-167.

Wittgenstein, L. (1974). *On certainty* (G. E. M. Anscombe & G. H. Von Wright, Eds.). Oxford, UK: Blackwell.

点估计(Point Estimate)

点估计是通过样本【Sample】对一个总体参数【Paramefer】作出的估计值。点估计是很有用的汇总统计量,在定量【Quantitive】分析报告中无处不在。点估计的实例包括回归【Regression】分析中获得的这类系数值,如截距【Intercept】和斜率【Slope】。点估计与区间估计(interval estimate)形成对比,后者又被称为置信区间【Confidence Interval】。置信区间反映了点估计的不确定性或误差范围。一个点估计提供了总体参数的单个值,而置信区间提供了一系列总体参数的可能值。虽然真正的参数是未知的,但点估计是最佳近似。

点估计便于变量【Variable】之间关系的描述。这种关系可能有统计显著性【Statistical Significance】,但这种关系的大小也是我们感兴趣的。点估计能提供被检验关系的更多信息。这里提供一个假设的简单回归分析以说明点估计的概念。给定自变量【Independent Variable】X 和因变量【Dependent Variable】Y 的回归方程是 $Y=a+bX+e$,令 X 等于考试之前夜间睡眠的小时数,Y 等于考试成绩(以百分比的形式),b 为斜率,a 为 y 轴的截距,e 为误差。假设数据的分析产生下列回归方程:$\hat{Y}=-2+9.4X$。点估计是从样本统计量中获得的值;在这里,我们感兴趣的是截距和斜率值。斜率表明,平均而言,额外增加 1 个小时的睡眠与考试成绩增加 9.4 个百分点是联系在一起的。如果斜率估计也是显著的,那么点估计表明,学生考试前一晚睡眠时间越长,则考试表现更好。另一方面,如果斜率估计是非常小的,比如说 0.5——那么两者的关联性就不会显得有实质性意义,尽管具有统计显著性。这个例子说明点估计提供了有关被检验关系的有价值的信息。

截距的值也是我们感兴趣的。对于这个例子,截距的值是-2。截距表明,平均来说,学生睡眠时间为 0 小时的成绩得分为-2 个百分点;但是,该值没有意义,因为考试得分范围为 0~100。而且,如果我们要沿 x 轴和 y 轴绘制这些数据点,我们会看到,没有或只有几个学生在测试前一夜睡得很少(0~2 小时)。由于数据点如此之少,应当避免对这些睡眠时间小的数据点的考试成绩进行预测【Prediction】(在这种情况下,我们得到的结果毫无意义)。点估计对于说明更多具体条件下的关系和进行预测是非常有用的,但是基于相对较小的数据所得到的点估计会包含更多的误差。在这种情况下,y 轴截距的点估计必须谨慎使用。

这里给出的假设的例子,是采用最小二乘【Ordinary Least Squares】估计【Estimation】量来说明点估计的概念;然而,参数的点估计也可使用其他估计方法,如最大似然法【Maximum Likelihood】。实际上,点估计,或者说由样本统计量获得的数值,部分地取决于研究者选择的估计量。因此,对研究问题给定的特定属性选择适当或最佳估计量对研究者来说是很重要的(Kennedy,1998,pp.5-6)。点估计也对模型的设定【Specification】、测量误差、样本容量和样本【Sample】抽取方式比较敏感。

——Jacque L. Amoureux

(叶鹏飞译 高勇校)

参考文献

Kennedy, P. (1998). *A guide to econometrics*. Cambridge: MIT Press.

Lewis-Beck, M. S. (1980). *Applied regression: An*

introduction (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-022). Newbury Park, CA: Sage.

Lewis-Beck, M. S. (1995). *Data analysis: An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-103). Thousand Oaks, CA: Sage.

泊松分布(Poisson Distribution)

泊松分布是用来测量某个离散【Discrete】正变量在一段观察时期内取特定值的概率【Probability】。在一个给定时间内,我们在接近无限次的试验中观察"成功"的总数。这个总数没有上限。试验或事件之间相互独立并且不能同时发生。例如,在给定的1小时内,我们可能接到10次推销电话;在这种情况下,泊松变量取值为10。我们定义一个泊松随机变量的期望值【Expected Value】($E[Y]$)为 λ,或者下一个时期内预期的发生率。在计算了 λ 值后(通过以往时期得到的结果),也可以应用泊松概率密度函数【Probability Density Function】来判断 Y 取值为 k 的概率:

$$\Pr(Y=k)=\frac{e^{-\lambda}\lambda^{k}}{k!}$$

尽管泊松·西莫恩-德尼(Siméon-Denis Poisson,1791—1840)最初提出泊松分布是将其作为二项概率分布函数的近似方法,但后来学者们也从一系列命题中推导出泊松分布。

泊松分布与二项分布【Binomial Distribution】非常相似,二项分布概括了 n 次伯努利试验的结果。当试验次数(n)很大时,成功机会(p)很小,并且 $\lambda=np$ 大小适中,即在稀有事件的情况下,泊松导出函数可作为二项随机变量取特定值时其概率的一种逼近方式(推导参见 Ross,2002)。泊松分布的方差公式也遵循二项分布:$\text{Var}(Y)=np(1-p)\approx\lambda$,当 n 很大且 p 很小时。二项分布和泊松分布的关键差别是我们知道二项分布试验(n)的次数,但泊松分布的试验次数未知且趋向无穷大。

社会科学家已经开始将泊松分布作为概率分布独立地用于描述每年战争次数和给定人群内特定罕见行为的发生率等多种现象。我们来考虑一个给定年份中总统否决权的使用数量。这个例子满足更普遍的泊松推导假设(参见 King,1998)。在不同议案中否决票是相互独立的。同一时间内不可能产生两次否决。设定 λ 为3.5(前一年总统否决率的平均值),我们能够计算总统在接下来的一年中至少三次议案中投否决票的概率:

$$\begin{aligned}\Pr(k\geqslant3)&=1-P(k=0)-P(k=1)-P(k=2)\\&=1-e^{-3.5}-3.5e^{-3.5}-\frac{(3.5)^{2}}{2}e^{-3.5}\\&=1-10.625e^{-3.5}=0.864\ 1\end{aligned}$$

当改变间隔的长度时,我们需要对方程式进行调整,将 t(间隔的长度)设定为附加参数【Parameter】。在这个例子中,我们将要计算的是总统在接下来两年中至少在三个议案中投否决票的概率(图1):

$$\Pr(Y=k)=\frac{e^{-\lambda t}(\lambda t)^{k}}{k!}$$

$$\begin{aligned}\Pr(k\geqslant3)&=1-P(k=0)-P(k=1)-P(k=2)\\&=1-e^{-7}-7e^{-7}-\frac{(7)^{2}}{2}e^{-7}\\&=1-32.5e^{-7}=0.970\ 4\end{aligned}$$

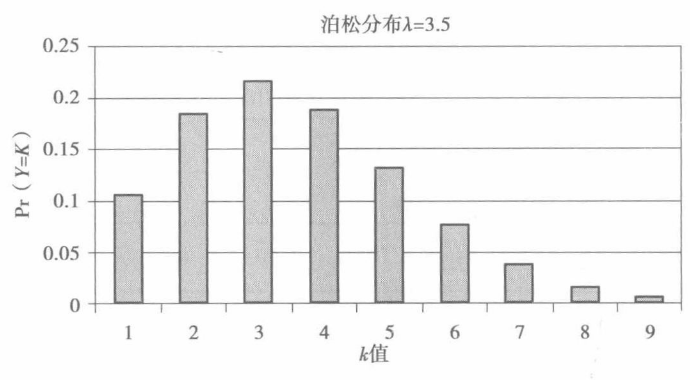


图 1　泊松分布

尽管上述泊松分布适用于单一观察值，但社会科学家已经发展出泊松回归【Poisson Regression】技术来对一组观察值的 λ 进行估计。

——Alison E. Post

（叶鹏飞译　高勇校）

参考文献

King, G. (1998). *Unifying political methodology*. Ann Arbor: University of Michigan Press.

Larsen, R. J., & Marx, M. L. (1990). *Statistics*. Englewood Cliffs, NJ: Prentice Hall.

Ross, S. (2002). *A first course in probability*. Upper Saddle River, NJ: Prentice Hall.

泊松回归(Poisson Regression)

泊松回归是一种回归模型【Regression Model】,其中由计数组成的因变量采用泊松分布【Poisson Distribution】的形式进行建模。计数因变量,只能取非负整数值,出现在许多社会科学情境中。例如,一个社会学家可能会对研究个体犯罪行为次数的影响因素感兴趣。在该例中,数据集中每个个体都与一个数值相联系,用来表明他或她在给定时间段内犯罪行为的次数。没有此类行为的个体计数为 0,有一次犯罪行为的个体计数为 1,以此类推。该问题中的社会学家可采用泊松回归来判断哪些给出的自变量(如个体受教育程度)与通过犯罪行为计数来测量的个体犯罪活动显著相关。

正如上面介绍的,泊松回归是以泊松分布为基础的。这种单参数一元分布【Distribution】支持非负整数,因此,在研究由计数组成的因变量时,适合使用这种分布。

模型设定

泊松回归的模型设定【Model Specification】基于如下假设,对于自变量向量 $\mathbf{x}_i$,计数因变量 y_i 对于每个观测 i 均独立服从泊松分布,其均值为 μ_i,且服从概率分布【Distribution】函数:

$$f(y_i \mid \mathbf{x}_i) = \frac{e^{-\mu_i}\mu_i^{\ y_i}}{y_i\,!} \tag{1}$$

μ_i 与 x_i 之间的关系被假定为:

$$\mu_i = \exp(\mathbf{x}_i'\boldsymbol{\beta}) \qquad (2)$$

其中，$\boldsymbol{\beta}$ 是待估计的参数向量（与 $\mathbf{x}_i$ 的维度相同）。注意 μ_i 和 $\mathbf{x}_i$ 之间的指数连接函数保证了 μ_i 只能是正数；这是很必要的，因为泊松分布的均值不可能是负的。基于一组 n 个观察值 $(y_i, \mathbf{x}_i)$，方程（1）和方程（2）联合产生对数似然函数

$$\log L = \sum_{i=1}^{n} [y_i \mathbf{x}_i'\boldsymbol{\beta} - \exp(\mathbf{x}_i'\boldsymbol{\beta}) - \log(y_i!)] \qquad (3)$$

其中，log（·）表示自然对数。这个对数似然函数的一阶条件——每一个元素 $\boldsymbol{\beta}$ 都有一个这种形式——是非线性的【Nonlinear】且没有闭合解。因此，必须应用数值方法来估计 $\hat{\boldsymbol{\beta}}$ 的值以使方程（3）最大化。

泊松回归结果的解释与其他非线性回归模型的解释类似。如果给定元素 $\hat{\boldsymbol{\beta}}$ 为正，那么 y_i 计数值会随着相应元素 $\mathbf{x}_i$ 的增加变大。然而，泊松回归模型中这种相对变化幅度计算起来很复杂，因为对于方程（2），μ_i 对某个 $\mathbf{x}_i$ 的导数依赖于 $\mathbf{x}_i$ 中的所有元素（对于不同元素 $\mathbf{x}_i$，μ_i 的交叉偏导数不为 0，这和在最小二乘【Ordinary Least Squares】回归中有所不同。）因此，对于估计值 $\hat{\boldsymbol{\beta}}$ 向量，可以改变 $\mathbf{x}_i$ 中的某个元素，以及在方程（2）中以 $\hat{\boldsymbol{\beta}}$ 代替 $\boldsymbol{\beta}$，基于改变的 $\mathbf{x}_i$ 向量产生一系列估计的 μ_i 值。然后应用方程（2）中的链接函数，直接产生一个非负整数概率分布——一个对于 y_i 的估计分布。需要指出的是，这一分布是以 $\mathbf{x}_i$ 中的其他元素不变为条件的。

方程（3）中的对数似然比（log likelihood）满足标准正则性条件，并且对 $\hat{\boldsymbol{\beta}}$ 的推导可采用标准渐近理论。因此，显著性检验、似然比检验等，可以很容易地根据 $\hat{\boldsymbol{\beta}}$ 和每个参数向量的估计方差矩阵来进行计算。泊松回归模型是一般线性模型【Generalized Linear Model】的特例（McCullagh & Nelder, 1989）；因此，一般线性模型理论可以应用于泊松回归。

扩展

标准泊松回归模型——方程（1）和方程（2）中概括的——可以被扩展成多种形式，其中很多形式由卡麦伦和特里维迪（Cameron & Trivedi, 1998）进行了阐述，其在计数回归方面非常详细并被广泛参考。泊松模型一个非常普遍的扩展是为了反映这一事实：计数因变量 y_i 在 $\mathbf{x}_i$ 给定时的方差往往比它服从泊松分布时本该有的方差值更大。注意均值为 λ 的泊松随机变量方差也是 λ。这意味着方程（2）中 y_i 的均值模拟同时也限定了 y_i 的方差。

对于给定观测值 $(y_i, \mathbf{x}_i)$，因为未观测到的异质性【Heterogeneity】，y_i 的方差超过 μ_i，于是 y_i 可以说产生过度离散（即产生比给定 y_i 和 $\mathbf{x}_i$ 期望更大的方差）。过度离散计数因变量可以使用负二项回归模型进行研究；标准泊松模型不应该被应用于过度离散的数据，因为基于这一模型的结果可能是误导性的。

此外，泊松回归模型是负二项模型的特例。因此，当应用负二项回归模型处理计数数据集时，它会直接检验以关联 $\mathbf{x}_i$ 向量中的自变量为条件的数据集是否是过度扩散的，如果是，就意味着不应该使用标准泊松模型进行分析。

标准泊松模型的另一个扩展归功于兰伯特（Lambert, 1992）的研究，其目的是模拟

含有过多零观测值的计数因变量(即大量的 y_i 值是 $y_i=0$)。这样的因变量数值的集合可以称为零膨胀,兰伯特的零膨胀泊松回归模型假设一些观测值 y_i 来源于标准泊松回归模型(或负二项模型),一些来源于"总是零"模型("always zero" model)。因此,每个观测值是0的 y_i,是来自"总是零"模型的潜在概率,以及相应地来自给零及所有正整数赋予正权重的常规泊松模型的潜在概率。标准泊松回归不是兰伯特零膨胀泊松模型的特例,且正如格林(Greene,2000)对于计数回归的讨论和研究中指出的,对于过多零值的检验需要非嵌套假设检验。

泊松回归模型的其他扩展可参见卡麦伦和特里维迪(Cameron & Trivedi,1998)以及格林(Greene,2000)的著作。以卡麦伦和特里维迪为例,他们讨论了泊松模型的时间序列【Time-Series】扩展以及推广,允许对计数向量进行建模。

——Michael C. Herron

(叶鹏飞译 高勇校)

参考文献

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data*. New York: Cambridge University Press.

Greene, W. H. (2000). *Econometric analysis* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Lambert, D. (1992). Zero-inflated Poisson regression with an application to defects in manufacturing. *Technometrics*, *34*, 1-14.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

政策导向研究(Policy-Oriented Research)

政策导向研究的目的是报告或者理解公众政策和社会政策过程的一个或多个方面,包括决策以及政策制定、实施和评估。需要区分的是为(for)政策的研究和对(of)政策的研究。为政策的研究是力图影响政策过程的各个阶段的研究(从政策形成之前直至政策的实施)。对政策的研究涉及问题如何界定、如何设置议程、政策如何制度化、如何作出决策,以及如何实施、评估和改变政策(Nutley & Webb, 2000, p.15)。政策导向研究可以同时是这两种类型。

政策导向研究在实际中如何影响政策(如果真是如此)是一个有争议的议题。扬、阿什比、博阿兹和格雷森(Young, Ashby, Boaz, & Grayson, 2002)确定了多种模式来帮助我们理解这一个过程。知识驱动模式中,研究(由专家开展的)引导政策。问题解决模式中,研究跟随政策,并且政策议题决定了研究重点。互动模式将研究和政策描述为相互的影响和作用。政治/策略模式认为政策是政治过程的结果,研究主题是政治驱动的。启蒙模式认为研究间接地服务于政策,讨论决策制定的情境,为理解和解释政策提供更宽广的框架。每一种模式都能帮助我们理解政策导向研究与特定政策发展之间的联系。扬与其同事(Young, Ashby, Boaz, & Grayson, 2002)认为,政策研究的角色"不只是解决问题,而是阐明议题并激发更广泛的公共讨论"(p.218)。

因此,政策导向研究可以具有多项功能,能拥有多种类型的受众。它提供一种专家功能,能够影响政策过程,并理解政策如

何运作以及针对决策者、政策网络和社区、熟知研究的实践者和学术界人士等目标受众来说怎样做才管用;它还提供一种民主功能,研究结论有助于推动开明的知识型社会发展和更广泛的民主进程。这里,研究的使用者将包括拥有既得利益的组织团体、服务使用者以及整个公众。

决策者和政策执行者(例如社会工作者、医护人员)越来越将他们的政策和实践建立在"证据"的基础上,特别是在有效性和"哪些有用"上的证据。将哪些视为证据(以及谁这样说)是存在高度争议的领域。虽然政策导向研究可以成为"证据"的一个关键来源,但其他来源也往往在实际中被当作证据使用,尤其是在影响政策和实践方面。

政策导向研究可以采用各种不同的研究设计和方法(Becker & Bryman, 2004)。有时候,根据要解决的特定研究问题只会应用一种方法。其他情况下,在一个单一研究或者作为跨地区或全国性大范围研究项目的一个部分,也可能会综合不同的方法。尽管每一种方法和设计都有自己的优势和局限,随机化控制试验【Randomized Controlled Trial】,然后是系统综述【Systematic Review】和统计元分析【Meta-Analysis】,都被很多人视为"黄金标准"。这在医疗卫生政策研究中特别明显,并且围绕哪些研究设计和方法最适合社会福利、教育、刑事司法以及其他公共和社会政策领域,存在持续的争论。

——Saul Becker

(叶鹏飞译 高勇校)

参考文献

Becker, S., & Bryman, A. (Eds.). (2004). *Understanding research for social policy and practice: Themes, methods and approaches*. Bristol, UK: Policy Press.

Nutley, S., & Webb, J. (2000). Evidence and the policy process. In H. Davies, S. Nutley, & P. Smith (Eds.), *What works? Evidence-based policy and practice in public services* (pp. 13-41). Bristol, UK: Policy Press.

Young, K., Ashby, D., Boaz, A., & Grayson, L. (2002). Social science and the evidence-based policy movement. *Social Policy and Society*, *1*(3), 215-224.

研究的政治(Politics of Research)

当研究涉及政治时,问题是研究是否应当追求或者能在多大程度上保持价值中立。早期的传统观点认为,社会研究应通过效法自然科学的实践和假定,力求科学性和客观性【Objectivity】,即哲学立场所称的实证主义【Positivism】。其关键特征在于运用价值中立的方法,作出合乎逻辑的论证,用观察到的社会事实来检验各种理念。这些原则一直饱受争议。但它们在19世纪和20世纪的大部分阶段都占据主流思想的地位。

最近,另一种观点开始流行。实证主义被人认为不适合于社会科学而受到挑战,社会科学关注的是社会行动者的意义、信念和价值观。对诸如调查【Survey】之类定量方法的依赖已经得到修正,更大的合法性被赋予深入的定性研究。社会研究与政策和实践之间的界限已经很模糊(Hammersley, 2000)。结果,社会科学能够产生客观知识的假定也已经受到挑战。

哈默斯利(Hammersley, 1995)提出,有

必要就社会研究是否是政治的提出三个问题。研究可能是非政治的吗？研究在实践中是否倾向于政治化？研究应该是政治性的吗？一些社会科学家认为，研究不可能不牵涉政治层面。所有研究必然会受到社会因素的影响，比如研究课题如何设计，怎样可能联系到参与者，以及研究过程中研究者的角色。资助者也在决定资助什么项目和与此相关的条件方面有重要影响。

还有人认为，社会科学家对社会世界采取漠然中立的态度既不可能也不必要。例如，社会学通常基于对“弱者”的关心，乐于向现状挑战。现在，女权主义和反种族主义的研究明确地以改变现实为目标。然而，仍然有人主张研究应该直接关注于创造知识而不是其他任何目标。根据这种观点，研究对于推动重大政治变革的能力是有限的，而且此类追求也限制了研究本身的价值（Hammersley，1995）。

走出这一难题的可能方法，是将知识创造过程中的政治与客观性区别开来。根据政治的或伦理的观点提出的问题，不能阻碍研究者在实际研究中对客观性的追求。这涉及将严格而系统的程序贯穿于研究全过程，包括分析期间和产生一种解释时。研究者们主张，即使并不存在社会科学家的绝对客观性的理想国，但仍然会有一种渴望去创造可靠的知识和理解（Christians，2000）。

——Mary Maynard

（叶鹏飞译　高勇校）

参考文献

Christians, C. (2000). Ethics and politics in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The handbook of qualitative research* (pp. 133-155). Thousand Oaks, CA: Sage.

Hammersley, M. (1995). *The politics of social research*. London: Sage.

Hammersley, M. (2000). *Taking sides in social research*. London: Routledge.

多边图（Polygon）

多边图是图表的一种类型，在 x 轴上显示一个定序、定距或分组比率尺度的变量，在 y 轴上表示另一个感兴趣的、通常是观察值频数的变量。

——Tim Futing Liao

（叶鹏飞译　高勇校）

＊也可参见频次分布图【Frequency Polygon】。

多项式方程（Polynomial Equation）

当方程右边包含幂次大于1的变量时，我们称之为多项式方程（polynomial equation）。在社会科学研究中，多项式方程较常见于经济学研究。例如，边际成本理论表明产量和边际成本之间呈U形关系。因此，当我们增加商品的产量，规模经济（因经营规模扩大而得到的经济节约）使得生产一个单位商品中的边际成本降低。然而，假如产量超出

一定数量，由于生产设备的饱和度，成本将趋向于增加。在数学上，这种关系可以用二次方程式表示：$\hat{Y}=\beta_0+\beta_1 X+\beta_2 X^2$。

多项式方程的随机形式

统计设定中，有时会用随机【Stochastic】形式来表达与多项式方程相关的理论。例如，上述方程的随机形式为：$Y=\beta_0+\beta_1 X+\beta_2 X^2+\varepsilon$。若扩展讨论至更一般的情况，若干变量为二次幂以上，则随机多项方程式的一般表达式如下：

$$Y=\beta_0+\sum_{i=1}^{m}\beta_i X^i+\cdots+\sum_{j=1}^{n}\zeta_j Z^j+\varepsilon$$

在本节的其余部分，我们将考虑随机环境下的多项方程式。

例如，回归【Regression】模型的多项表达式是非线性【Nonlinear】模型的特殊情况，它可以更一般地表示为：

$$\begin{aligned}Y&=\alpha g_1(\mathbf{z})+\beta g_2(\mathbf{z})+\cdots+\zeta g_k(\mathbf{z})+\varepsilon\\&=\alpha X_1+\beta X_2+\cdots+\zeta X_k+\varepsilon\end{aligned}$$

此外，$\mathbf{z}$ 是由若干自变量组成的向量。一系列转换【Transformation】函数，诸如 $g_1(\mathbf{z})$，$g_2(\mathbf{z})$，$g_k(\mathbf{z})$，指明了向量 $\mathbf{z}$ 中各变量之间的非线性关系。需要注意的是，我们以自变量的非线性表达开始，但通过自变量的转变，最终用经典的线性回归方程，即右边的类似表达为 $\alpha X_1+\beta X_2+\cdots+\zeta X_k+\varepsilon$ 来结束。这显示了模型的固有线性。如果我们发现非线性性质的关系，通常可将复杂模型转化为简单的线性形式。多项式方程是转换函数的可能表达之一。对数【Logarithm】、指数或倒数方程都是可选的，这取决于非线性的性质。

多项式回归的估计

我们可以用经典的最小二乘【Ordinary Least Squares】估计多项式回归。这是因为自变量的非线性不违反参数的线性假设。我们可以仅把自变量的转换值代入方程。然后，得到线性回归【Linear Regression】的简化形式。多项式回归中通常会存在多重共线性【Multicollinearity】问题，因为二次项和三次项的变量往往是相关的。然而，如果所有回归假设都满足，多项式回归方程的估计是最佳线性无偏估计量（Best Linear Unbiased Estimates，BLUE）。

多项式关系的检测

鉴于多项式回归方程的基本定义及其特点，一个自然的问题是，如何发现和确定多项式的关系。有两种方法可以回答这个问题。第一种是基于理论，第二种是基于统计学检验。基于现有研究的完备理论，比如前面提到的边际成本理论，我们可以给变量指定先验的关系，然后检验系数的统计显著性【Statistical Significance】。显然，这种显著性的检验是理论的自然检验。然而，在许多情况下，理论太薄弱不足以明确特定多项式或非线性形式。

因此，我们也可以从统计学上解决确定和检测多项式关系的问题。常用的方法有三种：

首先，可以用图形分析简单线性回归的残差。我们考虑只有一个变量的简单多项式方程。假设真实模型是$Y=\beta_0+\beta_1 X+\beta_2 X^2+\varepsilon$，但通过样本估计的回归方程为 $Y=\hat{\beta}_0+\hat{\beta}_1 X+\varepsilon$。因此，残差将反映 $\beta_2 X^2$ 和 ε。换言之，残差中除了真正的随机扰动之外，还包括真实模型中的系统变异。如果将观测的

x 值按升序重新排列,很容易就会发现残差的某种特征。

其次,可以采用分段线性回归。同样以残差分析为例,将观测值按升序重新排列,并将样本分为若干个亚组。如果对各个亚组分别进行回归,那么基于 β_2 的符号为正或负,得到的系数估计值会是持续增加或减少。

最后,如果在非线性的更一般的情况下来处理设定和检测问题,那么就可以使用 Box-Cox 设定。根据 Box-Cox 变换,可以明确自变量的非线性函数形式为:

$$X=\frac{Z^{\lambda}-1}{\lambda}$$

这里的 λ 是一个未知参数,所以对应不同的 λ 值,函数形式可能是多种多样的。我们目前兴趣的关键点是在 $\lambda=1$ 的情况下,X 是 Z 的线性函数。通常似然比【Likelihood Ratio】检验用于考察在有无 $\lambda=1$ 的限制情况下,其似然比统计量是否有差异。如果两者之间的差异有统计学意义,就拒绝变量符合线性的虚无假设。鉴于以上统计过程的结果,可以判断是否存在非线性。然而,非线性存在本身并没有告诉我们关于非线性的真实性质和形式。在非线性的一般设定中,可以运用 Box-Cox 变换公式来估算 λ 值。λ 值是非线性性质的标示。然而,Box-Cox 方法并不涵盖所有非线性关系的可能形式。此外,Box-Cox 方法并不只是针对基于多项式的非线性。

结论

综上所述,非线性的确定和检测,包括多项式的形式,应当基于理论,并通过统计学检验证实。纯理论的确定可能会弄错非线性的确切形式。单纯用统计学来确定非线性,则可能无法为非线性形式的运用提供坚实的基础。因此,明确多项式和其他非线性形式需要相当审慎的判断。

——B. Dan wood
Sung Ho Park
(叶鹏飞译　高勇校)

参考文献

Greene, W. H. (2003). *Econometric analysis* (5th ed.). Mahwah, NJ: Prentice Hall.

Gujarati, D. N. (1995). *Basic econometrics* (3rd ed.). New York: McGraw-Hill.

多分类变量(Polytomous Variable)

多分类变量拥有两个以上的不同类别,与二分【Dichotomous】变量形成对比。一般而言,任何包括两个以上观测值的连续【Continuous】变量、定距【Interval】变量或定比【Ratio】变量,都是多分类变量,但是,这个术语通常是用来表示定类【Nominal】和定序【Ordinal】变量,称为无序和有序多分类变量。当应用于定序变量时,该术语通常用于涉及具有相对较少的(如小于 10)不同类别的定序变量;拥有大量类别(如超过 20)的定序变量,有时被说成是连续的定距或定比变量,或者经常被当作与此相同。而且,尽管原则上此术语都适用于任何分析中的自变量【Independent Variable】或因变量【Dependent Variable】,但通常是用于 Logit 回归【Logit Regression】和 Logistic 回归【Logistic Regression】模型中的因变量。其他用于多分类变量的术语还有多分(只是另

外一种书写,也对应于二分变量)和多项(对应于二项)。因为二项和多项的概念既可以指特定概率分布,也可以指具有这些分布的变量,所以当指涉的并不是变量的分布时,可能更倾向于使用多分的概念。

——Scott Menard

(叶鹏飞译 高勇校)

合并调查(Pooled Surveys)

合并调查是指将多个不同的调查【Survey】合并到一个数据集。这些调查通常采用同一种测量工具,只是在不同时间重复实施。例如,分析人员可以结合或综合总统选举的历次调查,如1952—2000年,来进行美国国家选举的研究。这意味着在48年的时间段内,有一个由13次截面调查组成的集合。合并调查的优点是样本【Sample】量和方差【Variance】得到大幅度的增加。但由于将时间序列【Time-Series】和横截面设计【Cross-Sectional Designs】混合在一起,分析的复杂性也随之而来。而且,研究可能不得不局限在所有调查都涉及的范围狭窄的事项上。

——Michael S. Lewis-Beck

(叶鹏飞译 高勇校)

总体(Population)

总体是研究者要对其进行推论的分析单位的全集,在此分析单位可以是个人、社会群体、社会组织、社会产物等。在研究过程中,研究者会收集关于某一总体的可测量数据【Data】。在一项研究中,组成总体的分析单位被称为对象(subjects)。尽管研究的目标是对总体进行推论,但大多数情况下都不可能收集到研究者感兴趣的全部总体数据。因为诸如大学生或印第安人这样的总体都非常大,完整列举出所有案例是太费时间了,数据收集中也容易出现误差。收集总体的代表性子集通常比全部列举更可取,这种代表性子集被称为样本【Sample】。使用统计推论方法,研究者就可以基于总体的随机样本【Random Sample】而对总体特征进行推论。

有些社会研究关注的总体确实存在,如纽约大学的新生。但一项研究也可以关注概念性总体,它们在现实中并不存在,只是概念上假设出来的,如年龄在45岁及以上的离异的中产阶级男性。

总体的特征一般称为参数。惯例上,参数【Parameter】通常是用希腊字母表示,例如μ表示总体均值,σ表示总体标准差。总体参数是固定值,而且在大多数情况下是未知的。例如,犹他州支持堕胎的人口比例是未知的,但是如果几个研究者在某个时点测量这个参数,理论上讲他们会得出同一个数字。相反,不同样本中得到的样本统计量是不同的。用几个随机样本来测量支持堕胎合法化的犹他居民比例,几乎不可能得到同一个数字。不同于总体参数,某一样本的统计量取值是已知的,也可以计算出来。我们所不知道的是,这个样本在多大程度上能代表总体,或者说样本统计量能多准确地代表总体参数。

在进行实际研究之前,研究者通常要对欲推论的总体以及研究所使用的抽样方法作出假定【Assumptions】。这些假定通常可分为

两类:(a)研究者确信或愿意接受的假定;(b)要对之争论、需要经验证据的假定,因而也是研究者最感兴趣的假定。第一类假定称为模型【Model】,第二类假定称为假设【Hypothesis】。在研究过程中,对未知的总体参数进行假定后,研究者要会进行假设检验,以便确定如果这些假定实际上为真,得到目前的样本统计量的可能性有多大。

——Alexei Pavlichev

(李洋译 高勇校)

参考文献

Agresti, A., & Finlay, B. (1997). *Statistical methods for social sciences* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Babbie, E. R. (1995). *The practice of social research* (7th ed.).Belmont, CA: Wadsworth.

Blalock, H. M., Jr. (1979). *Social statistics* (Rev. 2nd ed.). New York: McGraw-Hill.

人口金字塔(Population Pyramid)

人口【Population】的年龄性别构成通常用一种称为人口金字塔的格式化图表示,从图1中可以看出按性别划分的分年龄组的人口绝对数量、百分比或者比例。这个称谓来自典型的三角形或金字塔形的人口分布图,这种样式表示人口在相当一段时期内保持了高生育水平而不断增长。如果人口增长非常缓慢或者规模稳定,则人口金字塔会呈蜂窝状;如果由于生育率很低而人口规模下降,人口金字塔会在底部显示出明显的收缩。

人口金字塔的优点在于能够清晰地展现年龄(或者出生年份)群组的规模,因此就能展现一些重要的有年龄限定的群体的相对规模,比如潜在的劳动力群体或诸如儿童和老年人的抚养人口群体。人口金字塔还可以显示出人口变动率(尤其是生育率)在当下和既往的趋势及波动对人口年龄结构的影响。由于人口金字塔可以同时显示出男女两种性别,因此从中也可以看出出生性别比的影响,在大多数国家中出生性别比是105名男性比100名女性,同时也能看出在死亡率和迁移率中性别差异的影响。人口金字塔有时也能显示分年龄统计的误差模式。国家统计机构在网上公布绘制人口金字塔所必需的年龄-性别统计表,这种做法已非常普遍。

图1的人口金字塔显示了1990年厄瓜多尔共和国人口的年龄性别结构,其中年龄用年来表示。从图中可以看出发展中国家的人口金字塔的典型特征。它的总体形状是三角形的,这是历史上人口生育率较高的结果。它在性别上并不对称:金字塔的底部男性略多,这是因为出生性别比偏于男性;顶部则女性略多,这是因为女性死亡率更低。有确凿证据表明在分年龄统计中存在"年龄堆积(age heaping)"形成的误差,这是年龄误报的常见模式,人们报出的年龄的尾数更可能是某些数字(通常是0和5,这里也是如此)而非其他数字(此类误差常常会被掩盖,因为人们通常会将5年作为时段来汇总年龄或出生年份)。金字塔的底部向内收缩,这里反映出来的要么是人口普查中有儿童漏报了,要么是厄瓜多尔在20世纪80年代生育率降低了。比较1990年人口普查中0~9岁年龄组的人口规模与2000年人口普查中10~19岁年龄组的人口规模,可以看出金字塔底部的收缩主要是由于1990年人口普查中有儿童漏报了。

——Gillian Stevens

(李洋译 高勇校)

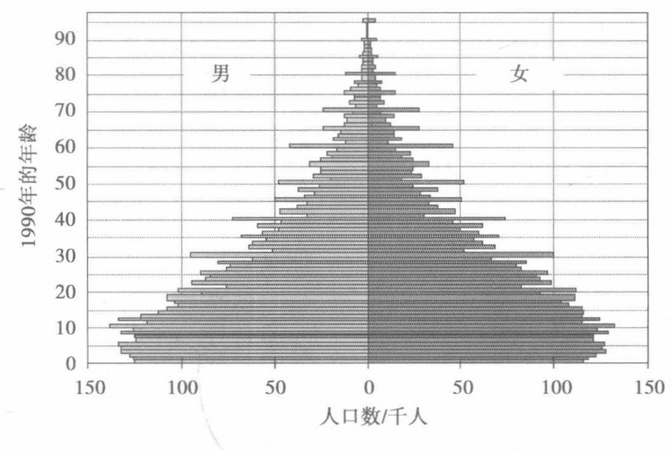


图 1 1990 年厄瓜多尔人口金字塔

（资料来源：Instituto Nacional de Estadistica y Censos, 1990.）

实证主义（Positivism）

实证主义是一种科学哲学，它拒斥形而上学的推测，主张用人的诸种感觉（senses）进行系统观察。此种观察如果能够有足够数量且彼此一致，就可认为得到了现象如何并存或相续发生的法则，对世界的"实证"知识就基于对观测的概推【Generalization】。在 19 世纪，实证主义不仅仅是一门科学哲学；它表达了崇尚科学成就的一种世界观。

中心原则

有很多人试图提出实证主义的中心原则，但如下方面被公认是其主要特征。

1.现象论。知识必须基于经验，必须基于观察者能够用诸种感觉感知到的东西。这种感知过程中不可掺杂认知（cognitive）过程的主观成分；它必须是意识清空后的"纯粹经验"。

2.唯名论。在科学解释中使用的任何抽象概念也必须来自经验；无法进行观测的形而上学概念并不是正当的存在，只不过是名称或字眼而已。因此，用来描述观察的语言不可被理论概念所影响。

3.原子论。经验（观察）的对象被认为是离散的、独立的、原子式的对事件的印象，它们构成了世界的最终的基本元素。这些原子式的印象可以用来进行概推，但概推指涉的并不是世界上的抽象对象，只是离散事件中存在的规则性。

4.普遍法则。科学理论被认为是一组高度概括的法则性表述；科学的目标就是建立这样的普遍法则。这些法则说明了现象之间存在着简单关系或固定关联，由此来对观察进行概括。解释【Explanation】就是把个体案例纳入合宜的法则之下。这些法则在范围上是普遍的，因为它涵盖了广泛的观察；这些法则在形式上是普适的，因为它可以无一例外地适用于不同时空。

5.价值判断和规范表述。"事实"和"价值"必须要分清，因为价值不是知识。价值表述没有可基于观察检验其正确性的经验内容。

6.验证。任何科学表述的对与错都可以用可观察的事件状态来评判。科学法则

可以用支持性证据的积累来验证。

7. 因果。实际上不存在因果关系【Causality】,只存在事件之间的规律性和恒定关联,如某一事件之后另一种事件相续发生。如果我们有的只是不同类型事件之间的规律性,那么解释只不过是把一个事件置于更广泛的规律性当中。

实证主义的各种类别

哈夫彭尼(Halfpenny,1982)认为实证主义可能有 12 种类别,而乌思怀特(Outhwaite,1987)认为可以精简为 3 大类。

第一种由奥古斯特·孔德(August Comte,1798—1857)提出,是不同于神学和形而上学的理解世界的方式。他认为所有科学形成了一个由彼此相关的不同层级构成的统一的等级结构,最基础的是数学,然后是天文学、物理、化学、生物学,最后是社会学。

第二种是逻辑实证主义,创立于 1920 年代的维也纳。这些哲学家们的招牌观点是,任何无法与某种事件状态对应(也就是说不能被经验证实)的概念或命题都是无意义的(现象论)。同时,他们认为较高层次科学的概念或命题可以化约为较低层次科学的概念或命题。换言之,他们持还原论立场,认为社会科学的命题最终可以还原为物理命题来分析。

第三种源于第二种,在科学哲学中有时被称为"标准观点",它在第二次世界大战后的英语世界中占主导地位。它的基本宗旨是,包括社会科学在内的所有科学都会重视以普适法则或普遍化的形式来进行解释。表明某一现象是某些法则的特例,这一现象就得到了解释。这些法则指涉的是事件之间存在的"恒定关联",或者在社会科学中,指涉的是统计相关或规则性("普遍法则")。

第三种实证主义相信对人和社会可以进行自然科学式的研究,这种信条也被称为科学方法的统一性(unity of scientific method)。它认为各种科学学科(包括自然科学和社会科学)虽然研究对象各有不同,但是可以采用相同的方法或解释逻辑,尽管每门科学都要以与其研究目标相适宜的方式来具体应用这些方法。

从最宽泛的意义上讲,实证主义是关于科学的本质、全能性及统一性的理论。在最极端的实证主义形式中,它认为只是科学知识才是有效的(非分析性)知识,认为此类知识存在于对可观察现象在空间上共存、在时间上相续的固定模式的描述中。它坚持科学的统一性,否定任何背离科学的知识,这就使它厌恶形而上学,坚持价值/事实的二分法,倾向于坚信科学指引下历史进步的必然性(Bhaskar,1986,p.226)。

社会科学中的实证主义

经过奥古斯特·孔德和埃米尔·涂尔干的努力,实证主义被引入社会科学中。各种形式的实证主义已经主导了社会学,尤其是在第二次世界大战后的数十年间,在诸如心理学和经济学的许多学科中的影响仍在继续上升。但是在最近实证主义遭受到了严厉的抨击(可以回顾 Blaikie, 1993, pp.101-104;Bryant,1985)。

批评意见

主要的几个争论点集中于实证主义的如下观点:经验是科学知识的合理基础,科学只能处理可观察的现象而不能处理抽象或者假设的存在,可以区分非理论的观察语言与理论的语言,理论概念与观察到的"现实"是一一对应的,科学法则是基于世界上

诸多事件之间的恒定联系,"事实"和"价值"可以割裂。

实证主义已经受到多方抨击,包括诠释主义【Interpretivism】、证伪主义【Falsificationism】以及批判实在论【Critical Realism】。诠释主义的批判重点是实证主义的本体论【Ontology】,认为实证主义对于社会现实的本质的观点是不充分的。诠释主义视社会建构的世界为社会现实,而实证主义则对此置之不论。实证主义者从社会世界对他们自己的意义出发建构了虚拟的社会世界,而忽略了社会世界对社会行动者的意义。

证伪主义的批判核心是实证主义"发现"知识的过程,以及为此知识辩护的基础。首先,证伪主义者认为经验作为知识的来源是不完备的,所有的经验都会牵涉诠释,故而他们认为不可能区分观察陈述和理论陈述;所有关于世界的陈述至少在一定程度上都是理论的。其次,他们认为经验作为评判知识有效的基础是不完备的,因为这导致了循环论证。除了再次诉诸经验以外,经验对于知识的评判又是建立在什么基础上的呢?

实证主义认为现实可以使用人的诸种感觉来直接感知,但这种观点已经被彻底驳倒。就算有一个独立于观察者的单独的、唯一的物质世界(这一假定并非无可置疑),观察它的过程中也必然会有意无意地进行诠释。观察是"饱含理论(theory laden)"的。无论是在日常生活中还是为了科学目的,人类观察周围世界的过程都不同于拍照。要"读出"哪些东西影响了我们的感受,我们不得不投身于一个复杂的过程,这个过程既要用到某一文化的语言中独有的概念,也会持有"那里"存在什么的期待。此外,我们在观察时并不是孤立的个体,而是文化群体或亚文化群体的成员,我们的本体论假定皆来自这些群体。因此,观察者是积极的主体,而非被动的容器。

观察和描述的本质皆是饱含理论的,对此实在论的解决方案是区分"与物相连(transitive)"的科学对象和"与物无涉(intransitive)"的科学对象。我们对经验领域的描述可能有赖于理论,但是实在领域的结构和机制独立于我们对它们的描述而存在。实在并不是实证主义认为的那样是被观察到的存在,也不只是一种社会建构;它就存在(it is just there)。因而,实在论者看来,在彼此竞争的各种理论中何者更成功地表现了这一现实,这是可以用理性判断来裁决的(见唯实【Realism】和批判实在论【Critical Realism】)。

批评实在论不仅认为实证主义的本体论是不完备的,它还批评了实证主义在事件之间建立固定关联的解释方法。即便看到两类现象有规律地同时出现,仍有需要回答为什么如此的问题。在批判实在论看来,确定可观察事件之间的规律性仅仅是科学发现过程的起点。事件之间的固定关联只会发生在实验条件下产生的封闭系统之中,而实证主义却将自然世界视为一个封闭系统。在像自然和社会的这种开放系统中,大量的生成机制会同时发挥作用、产生影响。因此,所观察到的"经验"关联并不能反映实际运行的机制的复杂性。

——Norman Blaikie

(李洋译　高勇校)

参考文献

Bhaskar, R. (1986). *Scientific realism and human emancipation*. London: Verso.

Blaikie, N. (1993). *Approaches to social enquiry*. Cambridge, UK: Polity.

Bryant, C. G. A. (1985). *Positivism in social theory*

and research. London: Macmillan.
Halfpenny, P. (1982). *Positivism and sociology: Explaining social life*. London: Allen & Unwin.
Outhwaite, W. (1987). *New philosophies of social science: Realism, hermeneutics and critical theory*. London: Macmillan.

事后比较(Post Hoc Comparison)

参见多重比较【Multiple Comparisons】。

后经验主义(Postempiricism)

很长时间以来,经验主义【Empiricism】一直是科学的同义词。毕竟,如果科学不能洞察现实,那什么能呢?在经验主义的历史中,约翰·洛克提出我们对世界的理解来自我们的经验。他同时强调分类必须基于对事物本质特性的了解。因而对事物自身的理解是没有特权通道的(privileged access),能直接了解的只有事物的属性(颜色、形状、感觉等)。经验主义哲学家大卫·休谟的观点更像一个怀疑论者,他认为我们只能了解外表,除此之外别无其他。

在一段复杂的历史中,经验主义为诸门科学的方法的统一性提供了基础。它还在社会科学和自然科学之间进行了严格划分,围绕的议题包括观察过程的中立性、经验无可争议的性质、资料【Data】和理论【Theory】的严格区分、达到知识基础的普适方法。托马斯·库恩、保罗·费耶阿本德和玛丽·赫西以其不同方式提出了后经验主义的观点,对上述理想主义假设提出了批评,以此将科学引向了更为社会化和历史性的理解,这种理解根源于解释【Explanations】和理解渗透于知识生产、发展和应用当中的各种条件。

库恩(Kuhn,1970)在《科学革命的结构》(*The Structure of Scientific Revolutions*)中为了简化其论证,认为科学是由多个"常规科学"时期组成的,此时科学家的工作就是在特定范式【Paradigm】的背景下进行"解谜(puzzle solving)"。范式由某一科学共同体的学术标准和实践构成,有着共同的理论假定。但是异常(anomalies)迟早会出现,关键理论将会被证伪。如果出现这种情况,重要的科学家就会开始挑战传统观点,危机就会发生。然后这一状况会继续扩散,最终发生科学革命,一种新的范式随之建立,然后以此类推。

从库恩(Kuhn,1970)的开创性著作以及后经验主义者的著作中,可以得到诸多推论。首先,逻辑实证主义【Logical Positivism】者认为必须将理论语言和观察【Observation】语言严格区分开来。卡尔·波普尔(Popper,1959)已经对此提出过批评,库恩的批评则更进一步。社会科学和自然科学之间的区分不能再简单地认为是事实和价值之间的区分。后经验主义者的反本质主义,促使他们致力于基于生产的社会情景进行详尽分析,这种生产是植根于特定环境中的实践活动。因此,社会活动并不是一个外在于科学探索的活动,其本质上就是科学探索的一部分。保罗·费耶阿本德(Feyerabend,1978)比持保守倾向的库恩更进一步,不仅认为局部因素决定了科学如何

进展,而且因为这一点,没有理由认为科学实践会有任何规则。

后经验主义批判开启了科学社会研究的一个丰富的传统。它也遗留下了一个核心问题,詹姆士·博曼(Bohman,1991)将之简洁地表述为:“在不重新落入非历史的、本质主义的认识论前提下,如何找到标准来比较和评估各种竞争性解释和研究计划。”(p.4)

——Tim May

(李洋译　高勇校)

参考文献

Bohman, J. (1991). *New philosophy of social science: Problems of indeterminacy*. Cambridge, UK: Polity.

Feyerabend, P. (1978). *Against method*. London: Verso.

Kuhn, T. (1970). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Popper, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson.

Williams, M., & May, T. (1996). *Introduction to the philosophy of social research*. London: Routledge Kegan Paul.

后验分布(Posterior Distribution)

后验分布 $p(\theta|\mathbf{x})$ 是用来描述人们在观察一个数据集 $\mathbf{x}$ 后对未知量 θ 的主观信念的概率分布。后验分布是使用贝叶斯规则来构建的,并且在贝叶斯推理【Bayesian Inference】中占据核心地位。将先验分布表示为 $p(\theta)$,似然函数表示为 $p(\mathbf{x}|\theta)$。贝叶斯规则通常写为:

$$p(\theta|\mathbf{x}) \propto p(\mathbf{x}|\theta)p(\theta)$$

贝叶斯规则描述了从数据中学习的过程。它解释了两个有着不同先验信念的个人在观察到 $\mathbf{x}$ 后为何会得出相似的结论。$p(\theta|\mathbf{x})$ 是一个概率分布,因此它在 θ 的所有可能取值上的积分必然等于 1,由此贝叶斯定理中的成比例性(proportionality)也可以得到解决。除了少数特例外,比例常数通常很难甚至不可能计算出来。在很难进行分析计算时,经常可用蒙特卡罗模拟方法(Monte Carlo simulation methods)来计算 $p(\theta|\mathbf{x})$ 的边缘分布或其他低维度的汇总分布。当 $p(\mathbf{x}|\theta)$ 和 $p(\theta)$ 是共轭分布时,可以很容易得到精确的分析计算值。如果似然函数和先验分布结合起来得到的后验函数具有与先验函数相同的函数形式,这一似然函数和先验分布就是共轭的。先验分布【Prior Distribution】的词条中列出了几对共轭的先验分布和似然函数。

例如,假设一人投掷一枚硬币 n 次,观测到 Y 次正面朝上,希望推断每次投掷硬币出现正面朝上的概率 θ。似然函数是二项似然:

$$p(Y|\theta) = \binom{n}{Y}\theta^{Y}(1-\theta)^{n-Y}$$

如果假设先验分布为 $\beta(a,b)$:

$$p(\theta) = \frac{\theta^{a-1}(1-\theta)^{b-1}}{\beta(a,b)},\ \theta \in (0,1)$$

这与二项似然是共轭的,于是后验分布为:

$$p(\theta|Y) \propto \binom{n}{Y}\theta^{Y}(1-\theta)^{n-Y}\frac{\theta^{a-1}(1-\theta)^{b-1}}{\beta(a,b)}$$

$$\propto \theta^{Y+a-1}(1-\theta)^{n-Y+b-1}$$

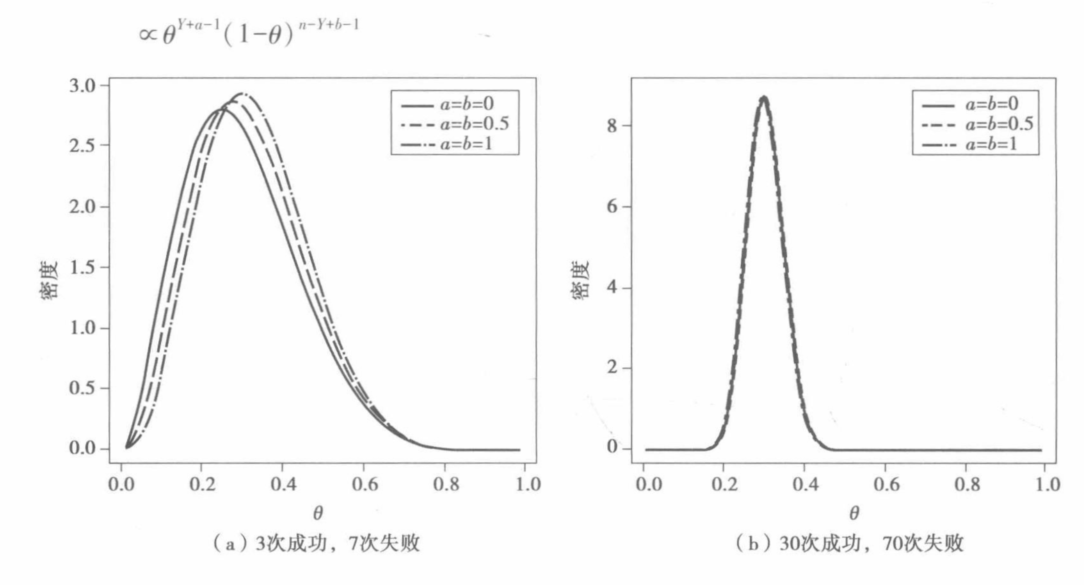


图 1 给定不同的成功和失败数以及不同 β 先验分布时 θ 的后验分布

后验分布与 $\beta(a+Y, b+n-Y)$ 分布是成比例的。由于先验分布是共轭的，就没有必要再进行复杂的积分计算了，因为后验分布族是已知的。只需要考察非标准化分布即可决定其参数。

β 分布的参数可以解释为成功和失败的数目。贝叶斯规则的修正只是将成功次数加到了 a 上面，将失败次数加到了 b 上面。将参数解释为先验的观测值，也是理解后验分布对先验分布的敏感度的一种方法。先验分布 $\beta(a,b)$ 会把 $a+b$ 个观测值的信息添加到数据中。贝叶斯分析需要先设定 a 和 b，但如果数据集中的成功与失败数目比 a 和 b 大得多，那么先验参数的选择对后验分布的影响就很小了。在小样本中，先验参数的选择则会显著地影响后验分布。

图 1 绘制的是给定不同的成功和失败数以及不同先验分布的 θ 的后验分布。这些分布可用来构造 θ 的后验均值、中位数或众数等点估计。θ 的区间估计（即置信区间）可通过确定相应 β 分布中 $\alpha/2$ 和 $1-\alpha/2$ 的分位点，或通过最高后验密度（Highest Posterior Density，HPD）的方法来获得。图 1(a) 中 θ 的后验分布展现出良好的品质，虽然它对于先验分布更加敏感。它是单峰的，被约束在 (0,1) 区间之内，偏度反映了成功和失败次数的不对称。对先验分布的敏感度通常意味着数据只能提供有限的信息，因此频率学派的方法会遇到困难。例如，对于图 1(a)，标准的频率学派方法的点估计为 $\hat{\theta}=0.3$，标准误差约为 0.15。如果 $\alpha<0.05$，那么频率学派近似得到的置信区间还会包括负值。

更多关于贝叶斯推理中后验分布的运用，请参阅 Gelman，Carlin，Stern 和 Rubin (1995)；Carlin 和 Louis (1998)。

——Steven L. Scott

（叶鹏飞译　高勇校）

参考文献

Carlin, B. P., & Louis, T. A. (1998). *Bayes and empirical Bayes methods for data analysis*. London: Chapman & Hall/CRC.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. London: Chapman & Hall.

后现代民族志(Postmodern Ethnography)

后现代民族志是在传统民族志的基础上对其反身性【Reflexivity】及其对所研究群体的责任进行了扩展的民族志【Ethnography】;它质疑传统民族志对真理性的宣称,并尝试采用新的报告模式。后现代民族志的新目标来自后现代主义。后现代主义【Postmodernism】包括一系列信条,涵盖从建筑学到文学批判主义、从人类学到社会学等多个学科。它拒绝宏大叙事和现代主义的元理论及其先入为主和一概而论的叙述。后现代主义提出要"打碎"这些叙述,研究其中的每个碎片。在社会学中,它主张研究日常生活的互动和细节,并基于其背景和脉络来单独分析每个部分。

反身性是后现代民族志的核心关注点,它认为长久以来民族志学者忽略了研究者的角色和权力。研究者会选择去观察什么,何时去观察,什么是重要的,什么才值得报告。在引用社会成员的话时,它们都会是剪裁出来的零星片段,用来支持研究者所提出的论证。但是,研究却忽略了这些问题,宣称自己具有中立性和客观性,用让"本地人"(研究对象)自己说话来隐藏这一切。后现代民族志尝试报告研究者的角色及其决策,以解决这些问题。这样尽管民族志学者仍然控制着选择过程,至少读者被告知了这个过程是如何进行的。此外,后现代民族志试图让"本地人"的声音能够更多地被听到。下面要讨论的各种报告模式都是为了这一目的。

责任(commitment)是后现代民族志中的另一个重要因素。传统民族志学者倾向于研究弱者和受压迫者,但是所获得的信息却很少用来减轻研究群体的痛苦。后现代民族志学者公开提出对研究对象的政治责任,许诺要分享他们的研究结果,以某种方式来帮助被研究者。因此,研究者不是为了自己的目的来控制研究和利用研究对象,而是与被研究者结为伙伴,共同致力于改善被研究者的境遇。

后现代民族志关注的第三个问题是对真理性的宣称。传统民族志倾向于采用"我在那里"的体例来书写(Geertz,1988;Van Maanen,1988)。也就是说,研究的合法性和可信性是通过研究者"亲眼目睹"来实现的。研究者对研究对象的关切进行总结和分类来讲述民族志。民族志学者会从田野笔记中摘取适当的示例性引语来加强自己的论题,从而制造出"本地人"自己说话的假象。后现代民族志质疑了与被访者越接近就离"真理"越近这种观点。实际上,它质疑绝对真理这一概念,认为我们能得到的只是相对的、暂时的视角,这种视角是受限于情境与历史的。我们在民族志中能够得到的只是对我们所见和被访者所讲的生活片段的理解。

认为民族志是相对的、有偏见的,这种信念影响了后现代的报告模式。后现代民族志学家不会依赖研究者(即作者)的权威,而是将其尽可能地降低。一些后现代民族志让许多研究对象表达他们自己的关注,尽量减少作者的评论(Krieger, 1983)。其他人则将民族志资料扩展到电影、电视、民谣、民间传说、绘画、梦及其他。一些人提出了自我民族志【Autoethnography】,用现在的眼光来审视过去,关注研究者对互动过程的感受和反应。一些人尝试用诗歌作为民族志的报告形式。戏剧和表演也成了后现代民族志流行的报告方式。此外,后现代民族志中还流行短故事的形式。所有这些形式都在尝试吸引更多听众,用更直接和引人入胜的方式来描述事件。

有一类后现代民族志是基于米歇尔·福柯的思想的。在福柯(Foucault,1985)看来,社会利用其权力来控制社会成员的方法就是为社会成员创造出用于理解现实的正规知识体系;例如,地图册是理解和描述苍穹的一种方式,牙科图表用来对我们口腔里的釉状物体进行命名和排序等。这些民族志学者的关注点是各种群体如何为其成员创造和控制现实。

一些女权主义者也在追求后现代民族志。她们拒绝男性主导下的传统民族志,批评传统的民族志显示了以男性为中心的立场和偏见(Clough,1998)。她们支持由女性研究女性,以摒弃男性主导的研究视角,从而允许女性自由表达自己。她们会拒斥她们所认为的男性研究人员的利用,倡导所研究的女性进行支持和合作。

最后一种后现代民族志是虚拟民族志【Virtual Ethnography】。一些研究者使用因特网来研究各种亚文化,他们会成为不同聊天室的成员或加入网络群组。这种民族志不再依赖面对面的交流,而是研究那些其存在只能通过电子媒介来理解的研究对象。

最后,后现代研究者还试图解决他们的另类民族志的信度【Reliability】和效度【Validity】问题。尽管尚未达成清晰的共识,但似乎已经出现了一系列基于美学信念和政治信念相结合的灵活原则。后现代民族志学者追求的范式连他们自己也承认有乌托邦色彩,这种范式的基础是认为写作并不是中立的而是政治性的,写作不仅要描述世界而且要改变世界。后现代民族志希望用一种参与的方式去描述世界;它想要改善所研究的不利群体的境遇,与此同时它还都想尽可能地忠实于所研究的个体对社会现实的看法。

——Andrea Fontana

(李洋译　高勇校)

参考文献

Clough, T. P. (1998). *The end(s) of ethnography: From realism to social criticism.* New York: Peter Lang.

Foucault, M. (1985). *Power/knowledge: Selected interviews and other writings: 1972-1977* (C. Gordon, Ed.). New York: Pantheon.

Geertz, C. (1988). *Works and lives: The anthropologist as author.* Stanford, CA: Stanford University Press.

Krieger, S. (1983). *The mirror's dance: Identity in a women's community.* Philadelphia: Temple University Press.

Markham, A. N. (1998). *Life online: Researching real experience in virtual space.* Walnut Creek, CA: AltaMira.

Van Maanen, J. (1988). *Tales of the field: On writing ethnography.* Chicago: University of Chicago Press.

后现代主义(Postmodernism)

后现代主义这一术语在诸多学术和文化领域中以相当不同的方式被人使用。使用同一词汇造成了一种令人困惑不解的、颇具误导性的印象,让人误以为它在社会、学术、文化领域的发展具有一致性和统一性。有三种讨论后现代主义的方式:作为建筑和艺术中的一种取向、作为描述当代(西方的或全球的)社会的一种方式,作为一种哲学。本文主要关注第三种意义上的后现代主义(也就是作为一种哲学或者社会科学中的一

种学术风格)，但是也有必要对其他用法进行简单回顾，因为各种用法通常被认为是相关联的。后现代主义很难去定义；定义后现代主义本身就违反了后现代主义的精神。但是，许多人用后现代主义这个词来指涉一种重视意义的碎片化及不稳定性的思想取向和文化取向：对知识、理性、社会制度、语言的剖析。

后现代主义一词在1970年代开始被人用来指称建筑中的一种从传统与本土中激发灵感的取向，这与非历史的、超理性的功能主义取向截然不同。后现代主义也意味着对不同建筑风格混合。在艺术领域，后现代主义既部分地继承了现代主义，同时又部分地突破了现代主义。艺术领域中的后现代主义摒弃了回溯推理(representation)的观念，强调的是挑战和新奇(如毕加索)，它实际上与后现代哲学有密切关系。后现代的理念是要在诸如高雅文化和低俗(大众)文化的不同领域之间“去差异化”，是要与艺术只局限于特定空间和特定功能的精英思想决裂，这可能构成了后现代主义的总体特征，这样一个总体标签似乎可以适用于多个领域中：

> 消除艺术和日常生活之间的界限；瓦解高雅文化和大众/流行文化之间的等级差异；创作风格上兼收并蓄、进行各种规则的混搭；进行恶搞、戏仿、嘲讽、戏谑来颂扬文化的“肤浅性”；艺术生产者的原创或天赋降低，认为艺术只能是重复而已。(Featherstone，1988，p.203)

有怀疑者认为后现代主义只在某些艺术批评家那里盛行，而没有在更大的艺术家群体中得到认同。但是如果说艺术领域中的这种趋势是更广泛的社会发展的表现，那么就可以认为后现代主义代表了一个新的社会或者社会发展的一个阶段。

后现代主义在1970年代晚期开始借助于法国和某些盎克鲁-撒克逊哲学家和社会科学家(如德里达、福柯、鲍德里亚、利奥塔、詹姆逊、哈威和鲍曼)进入社会科学中。其中有些人拒绝使用后现代主义这个词；福柯也反对“后现代社会或时期”这种提法。社会发展的一个新阶段的理念是在詹姆逊和哈威的影响下形成的，一定程度上利奥塔对此也有影响。

我们如果谈论到一个的社会阶段，就把后现代主义从文化领域拓展到了社会科学领域。许多学者谈论的是后现代性而不是后现代主义，我们在此也这样做。关于这一时期的意义，有好多种看法。有人认为后现代性意味着后工业主义和后资本主义，但是也有人认为后现代性是资本主义内部的文化变迁。这种阶段划分的观念表明这是有序列的：后现代主义(后现代性)是现代主义(现代性)之后出现的，但是人们仍在争论到底是何时出现的：有人认为是1920年代，有人认为是第二次世界大战之后，还有人认为是更近的时候。

一位有影响力的学者詹姆逊提到人们普遍感受在第二次世界大战之后一种新的社会开始出现(被描述为后工业社会、跨国资本主义、消费者社会、传媒社会等)。新类型的消费，有计划的淘汰；时尚和风格更加快速地转变；广告、电视和传媒以前所未有的程度渗透整个社会；城市和乡村、中央和外省之间的传统对立差异被取代为普遍而标准化的郊区生活方式；超级高速公路网络的发展和汽车文化的到来——这些特征都标志着新的社会与旧的战前社会完全不同了(Jameson，1983，pp.124-125)。

后现代性针对的是与启蒙运动相关的现代性思潮。在谈论现代性(以及后现代性)时，理性被视为指导原则和现代化方案中的可致目标，这是最常见的出发点。理性尤其体现在科学和知识提供的确定性和精

确性上，以及知识所带来的对自然的广泛控制和操纵。而后现代性代表的社会思潮和广泛的文化取向则通常质疑这种在科学、正规组织（科层体制）和社会工程学中所体现出的理性化方案的可能性和益处。“后现代社会”因此是反对现代性所宣称的确定性、客观性和发展取向的。

现代性和后现代性之间是有彼此相续的还是彼此重叠的，对此有不同看法。人们仍然争论后现代性的出现意味着对现代性的取代，还是意味着出现了一个与之并行且相悖的流派。对于很多评论者来说，后现代性最好应被视为对现代主义的一种回应或者评论，而非现代主义之后取代了现代主义的一个发展阶段。可以说，现代性的原则和实践在当代西方社会中仍占有主导地位。

后现代性有时是一种思考世界的基本方式，从而与后现代主义有所联系。有人认为，社会发展需要敏锐的词汇和思想，能对当下发展和现代社会发表有趣的论述。传统（现代主义的）社会科学及其语言被认为是过时的，无法表达精细而普遍的社会发展。但是反对意见认为，应当区分对后现代性的讨论和对后现代主义的讨论，以免混淆两者：社会趋势的时期划分和辨识是经验议题，而后现代主义的基础是哲学思想和学术风格的偏好，这和如何描述当代社会和社会变迁是两回事。什么时候谈论的是后现代性，什么时候作为后现代主义者谈论，这必须要清清楚楚。

谈完了后现代性以及历史分期和学术风格的关系之后，我们继续考察与传统社会理论和方法截然不同的框架和思想。在这个意义上，后现代主义与所谓的后结构主义高度重合，在学术议题和宗旨上两者基本上是同义词。但是后现代主义倾向于指涉的是社会或文化情境（如 Lyotard，1984），而后结构主义倾向于哲学，尤其是语言议题上。两种“主义”的渊源也有所差异，如利奥塔（Lyotard，1984）的后现代主义论述是从知识生产的条件的变化出生的，而后结构主义者则是对法国结构主义（如索绪尔）思想的背离。但是在社会科学中，后现代主义和后结构主义逐渐被视为同一个学术议题，只是后现代主义是一个更为流行的表达方式。

简而言之，后现代主义可以描述为“对统一性的抨击”（Power，1990）。对于后现代主义者来说，社会科学是一项粗陋和主观的事业，以试探性、碎片化和不确定性为特征。事实上，许多后现代主义者都不愿称其著作为社会科学。那些持极端后现代主义立场的人会贬低甚至拒斥社会与科学的部分。对后现代主义者来说，诸如系统、结构、能动者、共享意义、主导性社会秩序这些对社会的传统理解已经被传播过程中的话语（有强烈建构效应的语言）、图像和模仿所取代。

在社会科学中，后现代主义通常可以总结为如下五个假设和焦点（Alvesson & Deetz，2000；Smart，2000）：

- 话语的中心性（文本性）：强调语言的建构权力，“自然”对象被视为话语生成的。这意味着，没有对象或现象可以外在于语言或独立于语言。语言是形塑世界的活跃力量。不同话语意味着不同的社会现实。
- 碎片化的身份：强调主体性（subjectivity）是一个过程，个体的、自主、创造意义的主体消亡了。对个体的话语生产取代了对人的“本质主义的”传统理解。传统的“现代主义的”观点将人看作整合的整体，是意图、动机和意义的起源；与此不同，后现代主义者认为主体是去中心化的、碎片化的和受话语驱动的。例如，作为一个“女人”的意义与经验到底如何，对此没有一个绝对真实的答案，主导话语会生成“女性”

的不同侧面(如母亲、工人、选民、性感的个体),对不同时间的主体性有不同的影响。

● 对回溯推理(representation)观念的批判:语言具有不可判定性(indecidabilities),而不只是反映现实的镜子和传递意义的工具。径直地回溯推理客观现实是不可能的。不可能进行纯粹的描述,不可能将记录和分析分离开来。我们无法控制语言,但必须与语言的情境依赖性、隐喻性和建构性特性进行斗争(或者接受上述特性)。在社会科学中,围绕文本和作者身份的议题就变得(与在其他文化研究中一样)很重要。

● 放弃宏大叙事的基础和力量:强调多元发声(multiple voices)和局部政治,而不是理论框架和大规模的政治事业。不再有获取有效的普遍真理来指导长期、广泛的社会和科学事业的野心,知识和政治被视为都是局部的,由短期的、绩效导向的标准所决定的。有时更加重视多元发声和多元化。

● 权力与知识的关系:将权力与知识分离开来是不可能的,知识不是纯洁而中立的。权力和知识并不相同,但是他们彼此纠缠、互相依赖。知识总是具有权力效应;它能通过人们的接受来影响人们与其世界的关联方式,从而生成某一种现实并形塑现实。

严肃对待这五个观点,就会对社会研究的意义有彻底的重新认识。有人期待并希望"后现代转向"能够让社会科学更强调"社会现实的社会建构、自我身份的流动性而非固定性、真理的片面性,从而将替代那种客观现实的现代主义假定"(Lincoln & Guba,2000,p.178)。

后现代主义思想在多少与此不同的方向上被加以推进。大致可分为怀疑派后现代主义和肯定派后现代主义(Rosenau,1992)。怀疑派后现代主义主张一种"负面"议程,基本理念是不可能建立任何真理。回溯推理只是对事物强加一种武断意义的过程。研究成为一种解构,要去发现各种矛盾、被压抑的各种意义、潜藏在表面的可靠和可信之下的脆弱性,从而颠覆文本。这种方法不利于经验主义工作,至少是不利于传统的和实证的那种经验主义工作。

肯定派后现代主义同样质疑真理和效度的理念,但对社会研究持有积极态度。戏谑、嘲讽、幽默、拼搭、方法多元主义被加以赞赏。地方性知识(包括对情境知识的偏好和对寻求抽象普遍真理的拒斥)也被重视。这一派并不反对经验主义工作和经验主义立场,但是对描述、阐释、研究者权威等议题提出质疑。这已在社会和行为科学的许多领域引发了深入的方法争论,尤其是在人类学(Clifford & Marcus,1986)和女权主义(Nicholson,1990)领域中。

后现代主义可以看成是"语言学转向"的一个重要部分,"语言学转向"是哲学和社会科学中更广泛的一种趋势。这意味对语言进行彻底的重新认识:语言不是被动的意义传输媒介,不是社会生活和组织中无可置疑的部分,不只是能干的社会科学家熟练掌握的工具。相反,语言(话语)是社会研究者需要去研究的关键性的基础社会力量,社会研究者本身也被迫纠缠于回溯推理(representation)和作者身份等基础问题。

上述观点对于社会研究来说有着重要意义,但如前所述,这并不会终结经验社会科学而是会对经验社会科学有所启发并激发其活力(Alvesson,2002)。例如,访谈就不再被简单地视为对现实的反映或者被访者表达意义与经验的场合而已,访谈其实是

主导话语建构研究对象及其回答的局部环境。民族志不再被视为是基于详细田野工作的对其他文化的权威报告而是一种虚构文本,在这个文本中作者身份与研究对象的身份建构与要研究的就在“那里的”现象一样重要,甚至比其更重要。在后现代主义启发下的社会研究不是要捕捉意义、发现模式并得出确定的结论,而是要表现意义的不确定性,让多元声音发声,并向其他解读方式开放。

——Mats Alvesson

(李洋译 高勇校)

参考文献

Alvesson, M. (2002). *Postmodernism and social research*. Buckingham, UK: Open University Press.

Alvesson, M., & Deetz, S. (2000). *Doing critical management research*. London: Sage.

Clifford, J., & Marcus, G. E. (Eds.). (1986). *Writing culture*. Berkeley: University of California Press.

Featherstone, M. (1988). In pursuit of the postmodernism: An introduction. *Theory, Culture & Society*, 5, 195-215.

Jameson, F. (1983). Postmodernism and consumer society. In H. Foster (Ed.), *Postmodern culture*. London: Pluto.

Lincoln, Y., & Guba, E. (2000). Paradigmatic controversies, contradictions, and emerging confluences. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 163-188). Thousand Oaks, CA: Sage.

Lyotard, J.-F. (1984). *The postmodern condition: A report on knowledge*. Minneapolis: University of Minnesota Press.

Nicholson, L. (Ed.). (1990). *Feminism/postmodernism*. New York: Routledge Kegan Paul.

Power, M. (1990). Modernism, postmodernism and organisation. In J. Hassard & D. Pym (Eds.), *The theory and philosophy of organisations*. London: Routledge Kegan Paul.

Rosenau, P. M. (1992). *Post-modernism and the social sciences: Insights, inroads and intrusions*. Princeton, NJ: Princeton University Press.

Smart, B. (2000). Postmodem Social theory. In B. Turner (Ed.), *The Black well companion to social theory* (pp.447-480). Oxford. UK: Blackwell.

后结构主义(Poststructuralism)

第二次世界大战后的法国学术思潮不仅包括存在主义、现象学【Phenomenology】、马克思主义、心理分析,还包括索绪尔的结构主义著作。结构主义【Structuralism】与其他传统的区别在于它认为知识无法扎根于个体或者个体被历史给定的情境中。结构主义也反对经验主义,认为它无法区分一个现象的表象与其潜在“现实”。它是一种“关系的科学”,基于整体结构主义而强调社会中不同社会关系之间的相互联系或相互依赖。

被称为后结构主义的思潮反对结构主义在诸如列维-斯特劳斯的著作中体现出来的普遍主义追求。但是,与好多此类术语相同,后结构主义在界定了一种社会思潮的独特传统的同时,也具有使研究者降低敬业水平的副作用。

这个思想流派之下包括形形色色的思想家,而德国哲学家弗里德里希·尼采具有最显著的影响。他提倡视角主义

(perspectivism)的原则:也就是不存在用于审视真相的先验的优势视角,外部世界可以根据不同的信念和思想来进行诠释,而这些信念和思想都同样有效【Validity】。

后结构主义思潮的成员包括米歇尔·福柯、雅克·德里达、雅克·拉康、让-弗朗索瓦·利奥塔、吉尔·德勒兹、西克苏、露西·伊利格瑞以及朱丽娅·克里斯蒂娃。这一思潮在反人文主义和反本质主义的旗帜下批判性地审视公认的智慧,拒斥历史发展是理性的信念。随之而来的对主体一致性的批判,招致了强烈的批判性回应。要更好理解这一点及其对社会研究的意义,可以参看雅克·德里达和米歇尔·福柯这两位思想家的著作,他们都投入了这场争论中。

德里达的解构主义【Deconstructionism】的研究动力在于,要揭露西方哲学力图寻找透明主体的"梦想"。索绪尔认为,符号由"能指(Signifier)"(声音)和"所指(Signified)"(声音所指的思想或概念)两部分组成。通过语言这样一个自我指涉的系统,词汇的意义被确定了下来。但是德里达认为,我们看到一个符号时并不能马上明了其意义。不仅符号指涉的内容是缺失的,符号的意义也是缺失的。意义是在一系列能指中流动的,导致意义永远不会是固定的,也不会是透明的。无法找到一个固定点是基于延异(différance)的理念:也就是说,意义永远在逃避我们的理解,因为必须进行持续的澄清和界定。为此目的我们需要更多的语言,但是语言永远不能成为最终的裁决者。结果,符号只能用其他符号来研究,而其他符号的意义不断在逃避我们的理解。这使得德里达对那些希望通过将"先验的所指(transcendental signified)"掺杂其中而获得直接真理的人心怀警惕。

牢记这些批判,就可以对方法进行另类的解读。例如,科学可以解读为导致混淆的一系列尝试,而非通过心智觉悟而获得解放。价值与利益在原本认为其不存在之处被揭示出来,表明在原先假定中潜藏的是视角主义而绝非普遍主义。修辞将自身的预设潜藏于表面之下,而要识破这种伎俩就得在修辞学门下度过漫长的学徒期,我们也不能指望语言来解决这个问题,因为"不可裁决性(undecidability)"原则永远成立。质疑"在场的形而上学(the metaphysics of presence)"的理念影响了社会研究实践,例如在传统的对人际关系的研究会使用"意识性(consciousness)"和"意向性(intentionality)"来作为解释框架,而现在研究则关注社会世界是如何通过所谓"互文性(intertextuality)"而编织出来的。文本【Text】被视为社会实践,观察成为书写的行为,而非对独立现实的报道。

超越"意向性"来进行研究又陷入结构主义中去,这就把我们带到了福柯所提出的"阐释解析法(interpretive analytics)"的领域中。话语研究摒弃了表象(representations)的形成基础,摒弃了语言与现实之间达成透明关系的可能性,真理的政治交织起来构成了主体与客体。以这种方法来分析话语,就要秉持社会世界无法分成物质和精神两个领域的观念。话语既限制了经验的可能范围,也赋予了这些经验以意义。从这些角度来看,社会研究者不仅要考察话语的使用形式,也要考察话语的使用限度(后者在受福柯影响的研究中被忽视了)。

"事件化(eventalization)"的概念是福柯著作从考古学到谱系学构成的整体策略的一个部分,它认为一个事件的非凡突出(singularity)并不在于它在历史中的恒久或人类学眼中的不曾改变。福柯的方法是一种颠覆性的策略,它将看似不证自明的东西

打碎,表明“事情”何以可能是其他样子。一系列可能性就会随之涌现,而不必暗中认定历史之中有一个统一的主题或原因等待发现。社会研究者因此被要求去“发掘(rediscover)”可能性,而不是从事超越性的批判:“在某一时刻建立起来的关联、邂逅、支持、障碍、角力、策略等,随后会被认为是不证自明的、普遍的、必需的。在这个意义上,事实上你在使得多个诱因发生。”(Foucault,1991,p.76)

毋庸置疑,后结构主义提供了一个丰富的思想脉络,可以为社会生活提供新的洞见。同时,有些著作并不是总能够帮助社会研究者进行研究。这并不是说它不重要,只是说必须在其思想中寻求启发、澄清、契合的可能性。

——Tim May

(李洋译　高勇校)

参考文献

Foucault, M. (1991). Questions of method. In G. Burchell, C. Gordon, & P. Miller (Eds.), *The Foucault effect: Studies in governmentality* (pp. 73-86). London: Harvester Wheatsheaf.

Game, A. (1991). *Undoing the social: Towards a deconstructive sociology*. Buckingham, UK: Open University Press.

Kendall, G., & Wickham, G. (1999). *Using Foucault's methods*. London: Sage.

检验的统计效力(Power of a Test)

在对假设【Hypothesis】的显著性检验【Significance Testing】中,分析人员可能会犯两类错误:第一类错误【Type Ⅰ Error】和第二类错误【Type Ⅱ Error】。1 减去某一统计检验的第二类错误,就是这一检验的统计效力。检验的统计效力与真实的总体参数【Population Parameter】、与此检验有关的分布的总体方差【Variance】、样本【Sample】量、检验的显著性水平【Level of Significance】相关。

——Tim Futing Liao

(李洋译　高勇校)

幂转换(Power Transformations)

参见多项式方程【Polynomial Equation】。

实用主义(Pragmatism)

实用主义这个词的学术意义不同于其日常意义,指的是在美国 19 世纪下半叶出现,影响力一直延续到整个 20 世纪的一场哲学运动。实用主义的核心思想是任何概念的意义都是由其实践后果决定的;无论是在科学情境中还是在更广泛的生活中,任何

判断的真实性都要在实践活动中并通过实践活动来判定。

这些核心思想通常与其他一些思想立场有所关联。一个思想立场是拒斥许多理性主义者【Rationalism】和经验主义者【Empiricism】共同持有的一个假定,他们假定我们能够找到一些明确无疑的基础,然后基于此来建立知识。因此,对于实用主义的创建者皮尔斯来说,探究总是始于先前经验中产生的问题,总是要把多数经验视为当然:只去质疑当怀疑之处,而非所有可怀疑之处。同样,探究得来的知识也难免犯错;它并不是绝对无疑的正确。但是长期来看,探究是自我修正的:它不断发现和矫正错误来进行调整。

实证主义的另一个核心观点是,所有认知都是个体有机体与其环境之间持续进行的交易(transaction)的一个部分。这与古希腊颇有影响力的观点不一致,后者认为真正的知识来自与人类日常生活相脱离的旁观者。因此,另一个有影响力的实用主义哲学家威廉·詹姆斯认为,认知的功能不是完全复制现实,而是满足个体有机体的需求和利益,使之与其环境之间建立起生机勃勃的关系。对他来说,真理只是方便权宜之事。事实上,实用主义者认为哲学和科学从起源和功能上说也是实用主义的。因而,约翰·杜威认为科学探究过程是民主理念的核心,也是社会发展的适当基础。

除了共通之处外,实用主义传统内部也有分歧之处。有人认为科学才是理性思考的典范,也有人采用了更为包容性的路数。有人采纳唯实论【Realism】的立场,也有人认为我们只能接触事物在经验中的表象而非其后的现实。

在整个20世纪,很多有影响力的哲学家推动了实用主义思想的发展,著名的有刘易斯、奎因、帕特南、罗蒂和哈克。特别是罗蒂的研究引发了众多争议,尤其是它到底是一种真正的实用主义形式还是对之前传统的决裂。

实用主义对社会研究方法产生了相当大的影响力,尤其是通过社会学中的芝加哥学派,当然影响不限于此。卡普兰的《调查行为》(*The Conduct of inquiry*, Kaplan, 1964)受到实用主义的很大影响,是一本颇有价值的早期文献。

——Martyn Hammersley

(李洋译　高勇校)

参考文献

Haack, S. (1993). *Evidence and inquiry: Toward a reconstruction of epistemology*. Oxford, UK: Blackwell.

Kaplan, A. (1964). *The conduct of inquiry*. New York: Chandler.

Mounce, H. O. (1997). *The two pragmatisms: From Peirce to Rorty*. London: Routledge and Kegan Paul.

预编码(Precoding)

编码是数据文件中记录的一个数字(或字母或单词),它代表一项定量研究中某一被访者或回答人对某一封闭式问题【Close-Ended Question】的回答。一项有效的研究

在数据采集之时就对大部分信息进行编码，这被称为预编码。访谈员（或是自填问卷的受访者）会根据回答情况在相应数字上画圈，这个数字无须进一步加工或编码就可输入数据文件中。例如，图1显示了一个预编码的“就业状况”变量。

下面哪一项最能描述你当前的就业状况？
（请选择一项并画圈）
1.就业
2.自雇用
3.全日制教育/培训
4.失业
5.退休

图1 就业状况的预编码变量

预编码程序要有效，以下两点是必需的：

（1）类别必须明确清晰，彼此不能重叠。图1中的编码设计在这方面就有问题：一个人有可能同时就业而且接受全日制教育（如获得军方奖学金去上大学），或者同时就业而且接受全日制培训（如学徒工）。

（2）类别必须是穷尽的，覆盖全部范围。图1在这方面也存在问题：没有一个类别能涵盖全职家庭主妇。一些全职家庭主妇可能认为自己是“失业”；另一些可能会认为自己是“退休”。认为做家务和照顾孩子本身也是一种工作的那些人，则被迫选择上述两个选项，或者会误导性地选择“已就业”类别。换句话说，调查问题对社会世界的建构方式是某些被访者所不能接受的，这种错误在预编码的调查问题中并不少见。

解决这个问题的一种方法是提供一个“其他”选项，如果受访者状况无法归为任何一个选项时就可以选择“其他”项。这些回答将在事后不得不进行“办公室编码（office coded）”，或者将其划到一个现有选项中，或者另制订一个新的选项，或者归入一个包罗万象的“其他”选项。但是这会花费额外的时间和精力，所以如果可能尽量避免。

通常，我们采用单一编码变量，受访者被要求从一组编码中选择一个且只选择一个选项（图2）。可以进行多选的多编码变量（如“从下表中选出符合你情况的所有选项”）处理，但是这需要数据录入阶段进行特殊处理，以便分析时可以解释这些变量。

请圈出表示你最高受教育水平的数字。	
GCE或同等水平	1
“O”或同等水平	2
AS或同等水平	3
A+或同等水平	4
低于荣誉学士学位的高等教育	5
荣誉学士学位或同等水平	6
高于荣誉学士学位	7

图2 关于教育水平的一个问题

——Roger Sapsford

（李洋译 高勇校）

＊也可参见访谈表【Interview Schedule】、问卷【Questionnaire】。

先决变量（Predetermined Variable）

联立方程模型【Simultaneous Equations】中包含的外生变量和滞后内生变量【Endogenous Variable】通常被称为先决变量。变量的分类对模型识别来说至关重要，

因为在联立方程组中，自变量（或解释变量）中有很多是内生变量。

要介绍先决变量的概念，必须区分内生变量和外生变量。内生变量（或共同决定变量）的结果值要在模型之内决定。换句话说，内生变量的取值是通过在模型内与其他变量的相互作用来决定的。一个人对内生变量的行为感兴趣，是因为它们能被模型的其余部分解释。内生性（或联立性）术语反映了决定这些变量的模型的联立特征和反馈机制。

外生变量与内生变量的不同之处在于它们是为内生变量的行为提供解释。这些变量影响内生变量的结果值，但它们自己的取值完全是在模型之外决定的。从这个意义上讲，外生变量是表示原因的，塑造了决定内生变量的模型特征，但外生变量的取值并不相互【Reciprocal】影响。换句话说，不存在关于外生变量之间反馈互动或联立性的假设。

考虑一个非常简单的宏观经济模型，设定为有三个方程式的联立方程模型：

$$C_t=\alpha_0+\alpha_1 Y_t+\alpha_2 Y_{t-1}+\alpha_3 R_t+\varepsilon_{C_t} \quad (1a)$$

$$I_t=\beta_0+\beta_1 R_t+\varepsilon_{It} \quad (1b)$$

$$Y_t=C_t+I_t+G_t \quad (1c)$$

在该方程组中，同时期的消费（C_t）、投资（I_t）和收入（Y_t）是内生的，因为它们的数值是由其他变量解释的并在方程组内决定的。在内生变量之间存在反馈机制，因为在行为方程（1a）中，C_t 取决于 Y_t，而 Y_t 则通过恒等式（1c）取决于 C_t。利率（R_t）和政府支出（G_t）是外生变量，因为它们解释内生变量的行为，但它们的值是由方程组之外的因素决定的。可以看到这两个外生变量之间没有反馈机制。

在这个简单的宏观经济模型中，有一个滞后【Lag】内生变量 Y_{t-1}。这个滞后内生变量的值是由它过去的值来决定的。对于当前的收入值（Y_t），它的过去值（Y_{t-1}）是观察到的和先决的。因此，过去的内生变量可以放到外生变量的相同类别中。换句话说，滞后内生变量与外生变量是先决变量的两个子集。关于识别和估计联立方程的变量分类，更多的相关信息参考 Kmenta（1986）；Judge，Griffiths，Hill，Lütkepohl 和 Lee（1985）；Wooldridge（2003）。

——Cheng-Lung Wang

（叶鹏飞译　高勇校）

参考文献

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H., & Lee, T.-C. (1985). *The theory and practice of econometrics* (2nd ed.). New York: John Wiley.

Kmenta, J. (1986). *Elements of econometrics* (2nd ed.). New York: Macmillan.

Wooldridge, J. M. (2003). *Introductory econometrics: A modern approach* (2nd ed.). Cincinnati, OH: South-Western College Publishing.

预测（Prediction）

预测涉及从已知到未知事件的推理，无论过去、现在还是将来。在自然科学中，理

论常常先预测实体的存在，然后发现它。在社会科学中，预测通常是指关于未来结果的陈述。科学的预测不止是一个侥幸的猜测，而是需要经验证据的支持，并且必须以大量的理论为基础。因此，兰德尔·柯林斯（Collins，1995）基于地缘政治理论，证明自己成功地预测了东欧的巨变。

经济学家一般使用预测【Forecasting】胜于预测（prediction），二者有时是同义的。预测（forecasting）也被用在更严格的意义上，特指针对具体情况的预测，限定在一个既定的地点和时间范围：例如，预测未来 5 年中美国的汽车销售量。这个狭义上的预测（forecast）超出特定的情境是不能进行推广的。

预测的一种形式是外推【Extrapolation】，通过过去的趋势来设想未来的状况。为获得成功的预测，外推假设引发趋势的原因和社会过程是保持不变的。外推在短期内通常是准确的。然而，并不是所有的外推都可以被确定为预测。它们也许会成为可能发生的情形，告诫我们要改变行为以免发生不希望的结果。因此，关于全球变暖影响的外推，不一定是作为预测，而是作为劝勉，以减少对矿物燃料的消耗。

追溯涉及从结果出发反过来寻找原因，说明为什么一个已知的结果是可以预料到的。虽然追溯也许只是虚假的事后智慧，但追溯仍然可能是有效的，尽管预测是困难的或不可能的。铁木尔·库兰（Kuran，1995）指出，预测革命是否会发生的一个问题是伪造偏好（preference falsification），即人们会因为社会压力而歪曲自己的偏好。一个专制政权的公民，可能不愿意对一名社会研究者表达不满。在革命发生之后，预先存在不满的证据就变得很丰富。为了改善对革命的预测，我们可能因此需要分析标准社会调查之外的信息来源（如日记和回忆录、密函以及秘密档案）。

社会预测的一个关键问题是预测经常与他们预测的现象之间存在相互作用。在一个自我实现预言中，预测的实现恰恰是由于人们对于预测的解读；相反，一个自我否定预言的阐述确定了它将是错误的。一种实际的解决方案是将预测与可能对其作出反应的受众隔离开来。在某些情况下，宣称自我实现或自我否定的预言，可能是一种影响事件的有意尝试，以避免或减轻不良后果，或者增加预期结果的概率。

从某个系统的初始状态来预测其未来状态，要求初始状态中的微小变化对未来状态的影响是有限的。这在混沌系统（混沌理论【Chaos Theory】）中是不适用的，比如在气象学中，初始条件下无法觉察的差异会产生大相径庭的结果。

一个主要的困难是未来知识的必然不可预测性（Popper，1957）。我们无法理性地宣称拥有那些尚未被概念化或发现的全部预测知识。

——Alan Aldridge

（叶鹏飞译　高勇校）

参考文献

Collins，R.（1995）. Prediction in macrosociology：The case of the Soviet collapse. *American Journal of Sociology*，*100*（6），1552-1593.

Kuran，T.（1995）. The inevitability of future revolutionary surprises. *American Journal of Sociology*，*100*（6），1528-1551.

Popper，K. R.（1957）. *The poverty of historicism*. Boston：Beacon.

预测方程(Prediction Equation)

任何回归【Regression】类型的分析都提供两类信息:第一,它告诉我们一个或多个自变量【Independent Variable】对因变量【Dependent Variable】的估计效果。系数【Coefficient】和标准误差【Standard Error】估计值显示效果的相对强度及其估计精度。第二,回归类分析可以预测因变量的值。预测方程(prediction equation)涉及第二类信息。更具体地,预测方程是连接自变量与因变量预测值的等式。在线性回归【Linear Regression】分析中,预测方程是通过只取模型的系统部分来生成的,$\hat{Y}=\hat{\beta}_0+\hat{\beta}_1X_1+\cdots+\hat{\beta}_kX_k$,并用来计算因变量的预测值。在其他形式的分析中,如对数【Logit】、概率【Probit】及其他潜在指数模型,预测方程以非线性【Nonlinear】方式连接到可以预测的概率【Probability】函数。

一般地,样本内的平均预测误差是评估总体中模型拟合的实用标准。平均而言,有效的预测方程的预测值应接近样本因变量的实际值。然后,我们可以检验平均预测误差的显著性,以便将样本的模型拟合应用到总体。

预测方程也可以作为一种预测【Forecasting】工具,用于预测样本内(或外)的因变量个体值(individual values)。例如,假设我们要预测线性回归 $\hat{y}_i^0$ 的值,即样本内(或外)的单个观测值。那么我们会使用预测方程 $\hat{y}_i^0=\hat{\beta}_0+\hat{\beta}_1x_{1i}^0+\cdots+\hat{\beta}_kx_{ki}^0$。有效的预测方程应能准确地预测个体观测值,无论其位于样本内部或外部。因此,个体观测值的预测是评价模型质量的替代方法。

平均预测误差

在使用预测方程评估线性回归的平均预测误差时,有两种公认的标准:R^2 和估计标准误【Standard Error, SEE】(也可表示为 $\hat{\sigma}_e$)。R^2 统计量计算公式为:

$$R^2=1-\frac{\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}$$

注意,该计算需要预测误差的平方和(用预测方程计算),然后再与最简单的预测方法相比来对预测误差进行标准化。从1中减去这个量,产生一个0~1范围的拟合度量值。在度量的0值端,我们可以得出结论,该模型没有预测能力;在度量的1值端,我们可以得出结论,该模型可以无误地预测每一个观测。0和1之间的值是对平均预测能力的定量评估。

R^2 用于评估模型平均预测误差的缺点是,它不能显示以分析的测量单位表示的预测误差是多少。R^2 统计量取值范围为0~1,但对于中间数值,实际发生的预测误差大小是未知的。因此,一些分析者更喜欢使用估计标准误【Standard Error of Estimate】。其计算公式为:

$$SEE=\sqrt{\frac{\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2}{N-k}}$$

注意,SEE 计算需要预测误差方差的平方根,还是使用上述预测方程进行计算。结

果是按因变量的测量单位表示出的平均预测误差。SEE 统计量的数值,告诉分析者在使用上述预测方程时,预测的误差平均是多少(以因变量的度量单位表示)。这提供了对模型的预测能力更直观的描述。

从样本中计算出来的 R^2 或 SEE 也可以用来推断总体的平均预测误差,分析者通常计算 F 统计量,虚无假设为模型没有预测能力。对 R^2 而言,虚无假设为 R^2 统计量值等于零。对 SEE 而言,虚无假设为平均预测误差等于观测值与均值的平均偏差。这两个假设是等价的,从同一总体重复抽样时,分析人员可以借此推断预期平均预测误差的大小。

个体预测

用 R^2 和 SEE 衡量的平均预测误差,提供了对总体模型拟合的度量。然而,如果分析者想使用预测方程来预测因变量的特定值,则这些统计量就没法准确估计这时候的预测误差了。其中一个原因是靠近自变量分布末端的随机观测值变得越来越异常。朝向 X 分布的两端,数据变少,因而个体预测的不确定性更大。X 的观测值离典型值越远,个体预测误差就变得越大。例如,假设我们感兴趣的是通过预测方程 $\hat{y}_i^0 = \hat{\beta}_0 + \hat{\beta}_1 x_i^0$ 来预测因变量的个体观测值。使用该方程,个体观测的预测误差为:

$$\mathrm{SE}[y_i^0 - \hat{y}_i^0] = \mathrm{SEE}\sqrt{1 + \frac{1}{n} + \frac{(x_i^0 - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}$$

注意,这个统计量通过函数获得预测值距离分布中心的距离,对平均预测误差(前面讨论的 SEE)进行加权。此统计量的多元形式需要使用线性代数,参见 Greene (2003)。

如果我们想对多个个体进行预测,那么个体预测中的平均误差可以用来衡量预测准确率。这就是所谓的均方根【Root Mean Squared】误差(RMSE)或 Theil's U 统计量。该统计量表示为:

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i^0 - \hat{y}_i^0)^2}$$

其中,m 是个体预测的数量。注意,这正好是预测误差平方和的均值的平方根。RMSE 值越小,意味着预测方程的预测能力越强。

——B. Dan Wood

Sung Ho Park

(叶鹏飞译　高勇校)

参考文献

Greene, W. H. (2003). *Econometric analysis* (5th ed.). Mahwah, NJ: Prentice Hall.

King, G. (1990). Stochastic variation: A comment on Lewis-Beck and Skalaban's "The *R*-squared." *Political Analysis*, 2, 185-200.

Lewis-Beck, M. S., & Skalaban, A. (1990). The *R*-squared: Some straight talk. *Political Analysis*, 2, 153-171.

预测变量(Predictor Variable)

在每一个研究设计【Research Design】中,一系列自变量【Independent Variable】可以被认为是影响一个因变量【Dependent Variable】的前提条件。这些变量可以由研究人员实际操作得到(如在实验【Experiment】的情况下得到的数据),或仅由研究人员观察得到(如调查数据的二次分析【Secondary Analysis of Survey Data】)。如果是由研究人员简单观察得到的前提条件(即不进行实验干预),那么它通常被称为预测变量。

预测变量和自变量之间的区别要取决于具体研究设计。在实验性研究中,两者的意义是不同的,因为自变量被认为是在研究者的控制下(即它是实验中的操作或处理方式和手段),而预测变量被简单地视为一个前提条件(如对象的性别),从而判断其对因变量的预测情况。但是,在非实验研究(如相关性【Correlation】研究)中,因为研究人员没有操控模型【Model】中的任何变量,所以这种区分没有实际意义。

考虑下面两个研究设计:研究者感兴趣的是评估竞选广告对选民投票选择的影响。在第一种情况下,研究人员从选民的一个随机样本【Random Sampling】中收集以下变量的数据:(a)选民是否有可能把票投给候选人史密斯;(b)性别;(c)年龄;(d)政党认同;(e)是否接触过候选人史密斯的竞选广告。然后研究者利用这些数据,得出以下回归方程:

候选人史密斯所得选票=

$\beta_0+\beta_1$(性别)$+\beta_2$(年龄)$+\beta_3$(政党认同)$+\beta_4$(竞选广告)$+\varepsilon$ (1)

在第二种情况下,研究人员从一个选民的随机样本收集数据,只涉及(a)性别,(b)年龄,(c)政党认同三个变量。然后投票者被随机分为对照组和实验组。对实验组选民播放史密斯候选人的竞选广告,而对照组选民只观看产品广告。在对两组进行干预后(是否观看竞选广告),两组选民被问及他们是否可能会投票给候选人史密斯。以研究的结果为基础,研究人员给出(1)的回归方程。

通常,在第一个设计中,回归方程右边的所有变量被称为自变量(或回归量【Regressor】)。然而,在第二个设计中,我们可以对自变量和预测变量加以区分。在这种情况下,自变量是选民是否看到了候选人的广告。这是研究者最关心的关键的解释变量【Explanatory Variable】,也是由研究者进行干预的。这个模型中的预测变量是:(a)年龄,(b)性别,(3)政党认同。在研究中,每个对象在研究人员干预之前都对这些变量有一个取值(因此它们被正确地视为先行条件),这些取值被认为是对因变量上回答者得分的预测(因此命名为“预测”变量)。

因为预测变量和自变量之间的区别在数据的统计分析中是无关紧要的(注意,在上面的例子中,从第一种设计到第二种设计,在统计学分析技术上没有发生变化),所以在实践中经常忽略它。但是,这种区别对于研究的设计以及结果的诠释和解释都非常重要。

——Charles H. Ellis

(叶鹏飞译　高勇校)

参考文献

Frankfort-Nachmias, C., & Nachmias, D. (1996).

Research methods in the social sciences (5th ed.). New York: St. Martin's.

Howell, D. C. (1999). *Fundamental statistics for the behavioral sciences* (4th ed.). Belmont, CA: Duxbury.

Jaeger, R. M. (1993). *Statistics: A spectator sport* (2nd ed.). Newbury Park, CA: Sage.

试测(Pretest)

一个实地试测是一项调查【Survey】的预演。这些预研究【Pilot】是极其有用的工具,使研究者能够在实地研究之前,识别调查项目和/或数据收集协议【Protocol】中的潜在问题。尽管所有调查在启动之前都应当经过试验,但试测对于一个大型调查和/或复杂调查来说尤其是有价值的。它们在任何涉及计算机辅助数据收集【Computer-Assisted Data Collection】的调查中都是必不可少的。在这一调查阶段可能会发现一些潜在的重大失误并进行补救。

大致说来,一个试测是指运用研究设定的数据收集程序,从一个相对较小数量的被调查者【Respondent】中收集数据。理想的试测样本是从研究的抽样框【Sampling Frame】中进行选取的。当这种情况不符合实际或者不可行时,通常根据便利性和可得性来选择被调查者。不管怎样,试测样本越能反映研究样本的特征,其效果越好。因为研究需要访问员来实施,所以最好要有至少三名访问员参与试测问卷。

试测回答要制成表格,并且回答的边缘分布是检查潜在问题的指标,例如回答中的变差【Variation】太小(包括针对整体的变量以及子群)或不正确的跳转模式。对于那些很多被调查者表示不知道答案或者拒绝回答的项目,很可能需要进行修改和复测。

尽管试测可以用在各种调查模式中,无论是访问员实施(通过电话或当面)还是自填式(通过邮寄或电子方式),但是访问员介入的研究为研究者提供了考虑访问员与受访者之间互动的机会,并识别不能很好询问的问题。一旦访问员简要了解关于总体项目和具体问题的研究目标,他们就会在模拟大型调查要遵循的条件之下来与一个小规模演练样本的成员进行接触。

访谈在被调查者明确表示允许的情况下进行录音。随后,受过专业训练的编码人员收听录音带,将访问员与受访者之间的互动进行"行为编码"。对于每一个询问的问题,互动的双方都进行编码。对于访问员一方,注意问题是否以准确的措辞读出,对特定受访者不适用的项目是否正确地处理,以及是否进行了追问【Probing】和说明。对于受访者一方,记录受访者在问题完全读完之前打断访问员或者要求说明和解释的频率,以及他们提供的答案是否符合可供选择的选项。行为编码【Behavior Coding】可以对有问题的题项进行识别,即访问员发现很难按照标准化方式实施,以及问题或用词不能一致地为被调查者所理解。如果发现任何问题,这一工具或草案可以在实地调查之前进行修改。

——Patricia M. Gallagher

(叶鹏飞译　高勇校)

参考文献

Fowler, F. J., Jr. (1995). *Improving survey questions*. Thousand Oaks, CA: Sage.

Oksenberg, L., Cannell, C., &Kalton, G. (1991). New strategies for pretesting survey questions. *Journal of Official Statistics*, *7*(3), 349-365.

前测的敏感性(Pretest Sensitization)

当参与者在一项实验【Experiment】中接受前测时,会存在一种风险,即他们对前测要测量的变量是警惕的或敏感的,如此一来,他们的后测得分将会受到影响。换句话说,受前测敏感性影响的实验对象会对实验内容变得敏感,而这会影响他们的行为。所罗门四组实验【Solomon Four-Group Design】有时就被用来处理这一问题。

——Alan Bryman

(叶鹏飞译　高勇校)

启动(Priming)

个体在环境中的经验会临时地激活了一些心理概念。这些概念,包括特性、图式、态度、刻板印象、目标、心境、情绪和行为,在激活之后会更容易被想起。这时我们称概念被启动,即它们变得更有可能影响后续的思考、情感、判断和行为。启动也指涉那些被用来模拟在真实世界的经验中经常会发生激活的实验技术。

启动的术语是由卡尔·拉什利(Lashley,1951)首先用于指反应倾向的临时内部激活。启动过程开始时,某个心理概念通过在环境中的经历激活。被激活后的水平不可能马上回到它原来的水平;相反,它随着时间而消逝。这漫长的激活状态会影响随后的反应行为。因此,最近使用的概念反应更加的快速和简单,因为他们更容易被想起。通过这种机理,启动创造了一个内部的临时准备状态。更重要的是,启动是一个被动、认知的过程,不需要内部的动机和意图。

心理学研究中有几种常用的启动技巧(参见 Bargh & Chartrand, 2000)。最简单的也许是概念启动(conceptual priming)。在这项技巧中,研究者的第一任务是悄悄地激活一个概念。然后激活的概念很快会自动地影响个人的响应,随后的任务超出他们的意识和意图范围。社会心理学中的一项开创性实验运用概念启动展示出激活的概念能够影响未来的社会判断(Higgins, Rholes, & Jones, 1977)。参与者在一开始就被积极或消极的特质所启动。那些由积极特质所启动的人,比由消极特质所启动的人更能对行为模棱两可的目标个体形成良好的印象。概念启动有两种类型:在阈上启动(supraliminal priming)中,个体有意识地了解启动刺激,但并不知道启动刺激对其随后响应的影响;在阈下启动(subliminal priming)中,个体既不了解启动刺激,也不知道哪些刺激对后续反应的影响。

第二个启动技巧是延滞启动(carryover priming)。在这种类型的启动中,研究者首先让参与者审慎地和有意地利用启动的概念。这包括让参与者思考或写出目标、心态、记忆、人物或其他激发启动构想的主题。然后个体从事第二个、不相关的任务。在第一个任务中启动的构想,影响个人在第二个任务中的响应。重要的是,个人并没有意识到这种影响。通常情况下,延滞启动被用于当关键概念过于抽象或精细以至于无法进行概念启动时。

第三个启动技巧是顺序启动(sequential priming)。这种方法测试了不同精神表现之间的关联。研究者提供一个启动刺激,紧跟着是一个目标刺激。参与者必须以某种方式对目标作出响应(例如,拼出它的读音,确

定它是否是一个词,对它进行评价)。就启动促进对目标的响应而言,两个构想之间存在一种记忆上的联结。所有关于启动的研究最后必须要对启动的影响进行评估。

——Tanya L. Chartrand
Valerie E. Jefferis
(叶鹏飞译 高勇校)

参考文献

Bargh, J. A., & Chartrand, T. L. (2000). Mind in the middle: A practical guide to priming and automaticity research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 253-285). New York: Cambridge University Press.

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, *13*, 141-154.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior: The Hixon symposium* (pp. 112-136). New York: John Wiley.

主成分分析(Principal Components Analysis, PCA)

主成分分析(PCA)是探索性多元数据分析的主要工具,特别是当研究人员希望深入了解一组变量之间的关系以及评估与这些变量相关的个体时。这一基本技术是针对连续变量设计的,但目前已发展出新的变体,以满足定类和定序测量层次的变量,以及能够处理不同类型的测量层次相混合的变量。此外,该技术还可与其他技术如回归分析【Regression Analysis】结合使用。在本词条中,我们将集中于标准的主成分分析方法,而主成分分析在不同条件下的应用概览可参阅 Joliffe(1986)和 Jackson(1991),针对不同测量层次相混合变量的 PCA 陈述可参见 Meulman & Heiser(2000)。

理论

假设我们有 I 个个体在 J 个变量上的计分,并且这些变量之间的关系是,没有一个变量可以完全地被其他所有的变量预测到。于是这些变量构成了 J 维空间的各个轴,并且个体在这 J 个变量上的得分可以在此 J 维空间中进行描述。然而,观察一个高维的空间是不容易的;而且,个体在高维空间排列中的大多数变异可以在低维空间中进行展示而没有多大变异性的损失。例如,图 1 是样本 A 得分的二维椭圆图形,能够通过第一主成分在一个维度上很好地表示,并且只需沿着这一维度解释变异性。然而,对于样本 B 的得分,一个维度上的表示要差很多(即在样本 B 中由第一主成分所解释的方差要比个案 A 中低很多,由此解释了单一维度在样本 B 上是不充分的)。

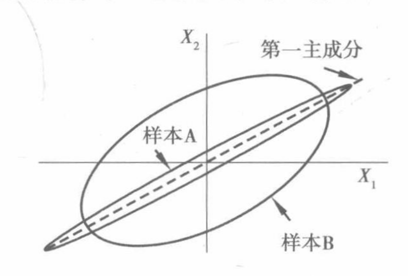


图 1 两个样本 A 和 B,在 X_1 和 X_2 变量上的得分

(注意:椭圆表示样本点的近似轮廓。第一主成分是样本 A 得分优良的一维近似值,但对于样本 B 并非如此。)

低维空间的坐标轴通常被称为成分。如果这些成分能够依次解释数据中的多数

变异性,那么它们就被称为主成分。个体在这些成分上的坐标被称为成分得分(component scores)。为了解释成分,还需得到变量在这些成分上的坐标,常见的方法是通过特征值【Eigenvalues】—特征向量【Eigenvector】技术。如果变量和成分都是标准化的,变量坐标即为变量与成分之间的相关性。通过检查这些通常被称为成分载荷(component loadings)的相关性,可以评估成分在多大程度上反映了某个变量的取值。特别是当一组变量与一个成分具有高度相关性,此成分就与所有这些变量具有共同点,并且在变量实质内容的基础上,可以试着去确定这些变量之间的共同要素可能是什么,并且假设这个成分就是在测量这个共同要素。

但是,主成分并不总是与一组特定的变量有着较高的相关性,因为主成分是用来顺次解释这些变量的最大方差,而不是产生最佳的解释。这也可以在由这些成分所决定的空间中对这些变量的绘图上看出(见图2)。这样的图形表明,在这个空间中还有其他方向的与成群变量相对应的成分,能够提供相对简单的解释。因此,主成分通常沿此方向进行旋转或变换。由于仅限于主成分的空间,所以这种变换并不影响这些成分对数据的联合拟合。当旋转导致相关的轴发生变化(斜交旋转,参见示例)时,这些变量的系数就不再是相关系数,并且通常被称为模式系数(pattern coefficients)。当(旋转的)成分互不相关时,每一个变量—成分相关的平方是由该成分解释变量方差的大小决定,而一个变量在所有成分上的相关平方和是用所得的成分预测该变量时的多重相关平方(通常称为公因子方差【Communality】)。

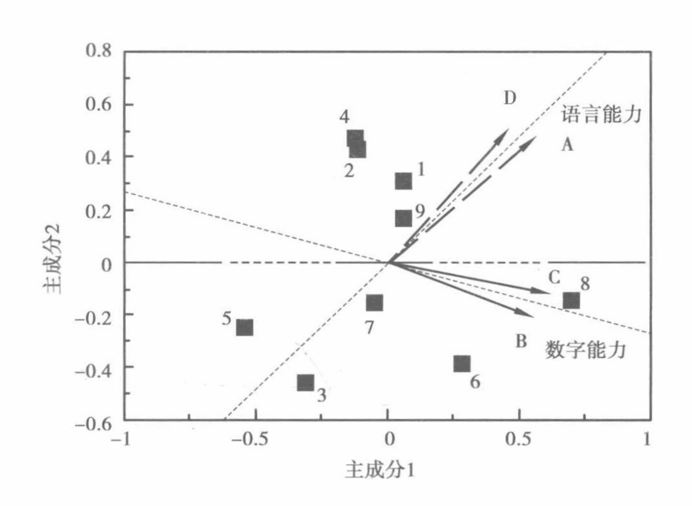


图 2 4 个测试子项和 9 个儿童的双标图示例,有两个主成分

(注:测试子项 A 和 D 测量数字能力,测试子项 B 和 D 测量语言能力。斜交旋转成分穿过每个测试子项组。)

任何成分分析的一个基本部分是通过一个双标图【Biplot】来联合描绘变量和样本得分之间的关系。就目测而言,这样的绘图最多是三维的,但是如果存在更多的维度,就需要构造几个图形(例如,为前两个维度绘制一个图形,再为第三和第四个维度进行绘图)。

举例

作为一个极其简化的例子,假设我们有 9 个孩子在包含 4 个测试子项的智力测验

上的得分，主成分分析表明，两个成分就足以对变异性进行充分描述。图 2(一个双标图)给出一个二维解的结果，描述了作为向量的测试子项 A,B,C 和 D,并且将孩子标识为从 1 至 9 的点。很明显，测试子项构成了两组，通过检查这两个分组可知，A 和 D 呈现的是语言智力，而 B 和 C 则是数字智力。子项与主成分并不完全对应，但是一个斜交变换使两个轴正好穿过两组的中间，因此它们可以被假设为分别代表数字智力和语言智力。旋转轴之间的角度表示它们是相关的，考虑到所有测试子项都是用来测量智力的，所以这并不奇怪。由于靠近数字智力轴的正向一侧，6 号和 8 号儿童在数字测试子项上的成绩特别好，而 3 号和 5 号孩子语言子项的成绩得分明显低于平均水平。由于 7 号孩子接近原点，所以他的得分接近所有子项成绩的平均分数（读图的规则，另请参阅双标图【Biplots】词条）。

——Pieter M. Kroonenberg

（叶鹏飞译　高勇校）

参考文献

Jackson, J. E. (1991). *A user's guide to principal components*. New York: John Wiley.

Joliffe, I. T. (1986). *Principal component analysis*. New York: Springer.

Meulman, J. J., & Heiser, W. J. (2000). *Categories*. Chicago: SPSS.

先验分布(Prior Distribution)

先验分布 $p(\theta)$ 是在对数据集 $\mathbf{x}$ 进行观察之前描述关于未知量 θ 的主观信念的概率分布。贝叶斯推论【Bayesian Inference】将先验分布与数据的似然性结合，形成后验分布【Posterior Distribution】用于对 θ 进行推断。本词条将从四个方面来描述贝叶斯推理中先验分布的作用：主观概率、客观贝叶斯、惩罚似然和层级模型。

基本的贝叶斯推理是一种用观察到的 $\mathbf{x}$ 更新关于 θ 的主观信念的理论。主观论者使用概率分布来描述个体的信念。研究者会考察个体在假设情境中愿意为 θ 的各个取值下注多少来得出每个个体的先验分布。这一过程得出的分布要适用于连续性参数空间，就需要设定模型。因为计算方便，经常使用共轭(conjugate)先验。如果先验分布和后验分布属于同一族，那么该模型就有一个共轭先验。表 1 列出了几个指数族成员的共轭似然先验关系。许多共轭似然先验族中的先验参数可能会被认为是先验数据，它提供了一种简单的方法来衡量先验分布的强度（例如，"某一个观察的先验信息的价值"）。使用共轭先验的计算实例参见后验分布【Posterior Distribution】词条。

表 1　几个普通似然的共轭先验分布

似　然	共轭先验
单变量正态	正态(均值参数)γ(gamma)(逆方差参数)
多变量正态	多元正态(均值向量)Wishart(逆方差矩阵)
二项式	β(Beta)
泊松/指数式	γ(gamma)
多项式	狄利克雷分布(Dirichlet)

任何较弱的先验对后验分布的影响是可以忽略的。然而也存在反例，此外，贝叶斯方法的批评者发现先验分布的获取存在

巨大困难(Efron, 1986)。客观贝叶斯(Objective Bayesians)通过"参照先验分布(reference priors)"尝试在标准情况下对先验未知进行建模来回应这种批评。卡斯和瓦塞尔曼(Kass & Wasserman,1996)在大量的文献基础上,对参照先验分布进行了研究。许多参照先验分布是积分反常的(improper)(也就是说它们积分趋于∞)。例如,如果 θ 的参数空间是无限的,均值参数的常见"无信息"先验分布("noninformative" prior)是均匀先验分布(uniform prior) $p(\theta)\propto 1$,这显然是积分反常的。只要后验分布是适当的,似然性得到足够好的表现,那么积分反常的先验并不会造成困难,这通常发生在频率论过程表现良好的简单问题中。然而,当复杂的模型中使用积分反常的先验分布时,后验分布的适当性就难以检查(Hobert & Casella,1996)。

参照先验分布的一个困难是,变量经过变换之后,单一尺度上的无信息先验分布变成了信息先验。例如,采用雅哥比介绍的对数变换,在 $\log \sigma^2$ 上的均匀先验分布变成 $p(\sigma^2)\propto 1/\sigma^2$。杰弗里先验是一个重要的参照先验族,它们不因变量变化而改变。模型 $p(\mathbf{x}|\theta)$ 的一般杰弗里先验分布为 $p(\theta)\propto \det(J(\theta))^{1/2}$,其中 $J(\theta)$ 是从单一观测中得到的费希尔(Fisher)信息。

虽然客观贝叶斯力求最小化先验的影响,惩罚似然的支持者仍然积极寻找可以改进高维模型中高变异的参数估计的先验分布(Hastie, Tibshirani, & Friedman, 2001)。例如,对于回归模型 $\mathbf{y}\sim N(\mathbf{X}\beta,\sigma^2 I_n)$,其中,$\mathbf{X}$ 是 $n\times p$ 设计矩阵,$\mathbf{y}$ 是 $n\times 1$ 响应向量。如果 $\mathbf{X}$ 不符合模型的假设条件,那么 β 的最小二乘估计是不可靠的。这种情况的出现,可能是由于数据集中存在明显的共线性,或者因为添加了多个交互作用项、多项式或其他类似样条函数的非线性基础函数而导致的。

由先验分布 $\beta\sim N(0,\tau^2 I_p)$ 产生等价于岭回归的 β 的后验分布:

$$p(\beta|\mathbf{X},\mathbf{y},\tau,\sigma^2)=N(B,\Omega^{-1})$$

后验分布的精度(逆方差) $\Omega=(\mathbf{X}^{\mathrm{T}}\mathbf{X}/\sigma^2+I_p/\tau^2)$ 是先验分布精度 $\Omega_0=I_p/\tau^2$ 与数据精度 $\Omega_1=\mathbf{X}^{\mathrm{T}}\mathbf{X}/\sigma^2$ 之和。如果 $\hat{\beta}$ 是 β 的最小二乘估计,那么后验分布的均值 B 可以表示为:

$$B=\Omega^{-1}\Omega_1\hat{\beta}$$

因此,B 是 β 和先验零均值的精度加权平均数。它是先验信息与似然信息之间的一个折中。先验分布确保 Ω 是正定的。当 $\mathbf{X}$ 不符合模型的假设条件时,先验分布起到稳定 Ω 逆矩阵的作用,这极大地减少了 β 的后验方差。

最后,先验分布可作为模拟复杂数据结构之间关系的方法。常被用来模拟嵌套数据的层级(非)线性模型【Hierarchical (non) Linear Models】就是一个很好的例子。层级模型认为在某一亚组的参数给定时,该亚组中的观测(例如,同一所学校中的学生考试成绩)是条件独立的。不同亚组的参数通过一个先验分布而连接起来,这个先验分布的参数被称为"超参数"。该先验分布允许亚组的参数向其他亚组"借力"来使得稀疏亚组的参数可以被准确估计。

举例说明,用 y_{ij} 表示学生 j 在学校 i 中的分数。假设同一所学校中的学生考试成绩在给定学校参数下是条件独立的,$y_{ij}\sim N(\theta_i,\sigma^2)$。假定存在先验分布 $\theta_i\sim N(\alpha,\tau^2)$ 由此可以得到同一所学校中的学生的后验相关性。例如,在上述前提下,

同一所学校中任何两个人的分数之间的相关系数是$\tau^2/(\sigma^2+\tau^2)$。图1提供了模型的图形化描述[著名的有向无环图(Directed Acyclic Graph, DAG)]。

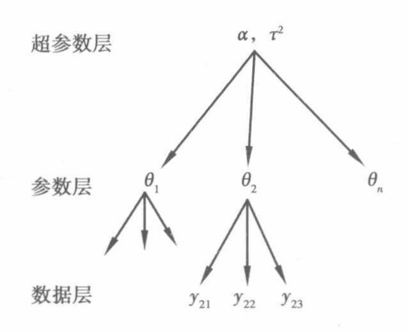


图1 层级模型的图形表示

(注:每个变量在给定上一级元素的情况下,与非下一级元素是条件独立的。)

若要查看先验分布如何允许一个亚组向总体子集"借力",要从数据层向上可视化(α,τ^2)的信息。然后,在给出了数据和所有其他参数的情况下,θ_i的后验均值为:

$$E(\theta_i|\mathbf{y},\alpha,\sigma^2,\tau^2)=\frac{\frac{n_i}{\sigma^2}\bar{y}_i+\frac{1}{\tau^2}\alpha}{\frac{n_i}{\sigma^2}+\frac{1}{\tau^2}}$$

由于基于数据的估计向先验分布收缩,所以此表达式被称为收缩估计(shrinkage estimate)。它结合了从全部样本中获得的信息以及依据(α,τ^2)汇总的特定子集i的数据信息。由τ^2测量的子集之间的变异,决定了收缩的数量。如果$\tau^2\to\infty$,那么可以从子集i的样本均值来估计θ_i,但如果$\tau^2\to 0$,那么所有子集都共享一个相同的均值α。θ_i的后验方差为:

$$\mathrm{Var}(\theta_i|\mathbf{y},\alpha,\tau^2,\sigma^2)=\left(\frac{n_i}{\sigma^2}+\frac{1}{\tau^2}\right)^{-1}$$

用语言来表述,后验精度是数据精度与先验精度之和。当n_i很小时,相比于独立估计每个θ_i,收缩估计可以大大减少θ_i的后验方差。关于层级模型的完整讨论,参阅Carlin & Louis(1998)。

——Steven L. Scott

(叶鹏飞译　高勇校)

参考文献

Carlin, B. P., & Louis, T. A. (1998). *Bayes and empirical Bayes methods for data analysis*. London: Chapman & Hall/CRC.

Efron, B. (1986). Why isn't everyone a Bayesian? *The American Statistician*, *40*, 1-5.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning*. New York: Springer.

Hobert, J. P., & Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, *91*, 1461-1473.

Kass, R. E., & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, *91*, 1343-1370.

先验概率(Prior Probability)

所谓先验概率,是这样一种概率【Probabilisty】的推理,即利用观察数据来更新在数据收集之前存在的概率。这种方法被称为贝叶斯推理【Bayesian Inference】,它与现有的其他方法即频率论形成了鲜明的对比。在频率论方法中,概率本身被认为是固定的但也是未知的量,可通过相对频次【Relative Frequencies】进行观测。如果将贝

叶斯方法应用到某一事件,假定研究者对该事件的概率(先验概率)有一些了解,运用贝叶斯定理【Bayes Theorem】,就能转化成后验概率。先验概率表示在实际观察资料完全得到之前就可以获得的知识,而后验概率则综合了先验概率和数据中的信息。这一变更过程运用了贝叶斯因子【Bayes Factor】。

先验概率相当于平常所说的概率,而且只有在我们的目标是运用新数据对它们进行更新时,才将其称为先验概率。当先验概率来自普查或一个早期的大型调查【Survey】,我们要基于近期的较小调查对其更新时,就会出现这种情况。虽然在大多数情况下,普查数据被认为比调查数据更可靠,但也会变得过时,而新的调查数据则能反映普查实施以来的社会变化。在这种情况下,通过考虑从普查资料中计算出来的作为先验概率的某一事实的概率,并运用最新调查资料对它进行更新,就可实现新旧信息片段的结合。当先验概率来自一个大型调查时,类似的情况也会发生。

例如,假设根据完成再培训项目的失业者在一段时间内找到工作的机会(概率)来研究项目的效率。当项目开始后,详细地监测项目效率,数据被用来确定不同组失业者的再就业概率。在第一年实施之后,为降低成本,停止了详细的监测,仅使用小规模调查来提供项目效率的数据。在该案例中,通过完整监测得到的概率要比通过后续调查计算出来的概率更加可靠。另一方面,在此期间,项目效率可能会因为一些原因而改变,并且调查结果可能会反映出这种变化。因此,有必要将原先的概率作为先验概率,并利用新的数据来对它们进行更新。

最简单的更新过程是使用所谓的贝叶斯定理,还可用来更新不可能同时发生的某些事件(如不同的宗教归属)的先验概率,并且可以适合进一步发生的事件(如同意政教分离的某种陈述)。如果前者事件的(先验)概率(如来自普查数据)已知,并且属于某一种宗教信仰的人同意政教分离陈述的条件概率是已知的,那么如果一位被调查者被观察到同意陈述,则对于该调查者而言,不同宗教归属的概率也可以被更新而产生新的概率。

——Tamás Rudas

(叶鹏飞译　高勇校)

参考文献

Iversen, G. R. (1984). *Bayesian statistical inference* (Quantitative Applications in the Social Sciences, 07-043). Beverly Hills, CA: Sage.

囚徒困境(Prisoner's Dilemma,PD)

囚徒困境是用来研究人们彼此追求个人利益时的动机与行为的一种博弈。在一个典型的囚徒困境博弈中,两个玩家各自在两种行为之间选择:合作与背叛。每个玩家得到的收益不仅取决于他自己的选择,还取决于另一个玩家的选择。在他们都作出选择之前,谁也不知道另一位选择了什么。收益是结构化的,所以无论另一位玩家怎样做,选择背叛总是可能会对选择者产生最高的回报;但是,如果两个玩家都选择背叛,每个人得到的收益要少于两人合作的情况——因此出现困境。

表 1 囚徒困境中的典型收益

		玩家 2 的选择	
		C_2	D_2
玩家 1 的选择	C_1	3,3	0,5
	D_1	5,0	1,1

资料来源:改编自 Rapoport & Chammah (1965)。

(注:每个单元格中第一个数字是玩家 1 的收益;第二个数字是玩家 2 的收益。C_1 和 C_2 分别表示玩家 1 和玩家 2 作出的合作选择;D_1 和 D_2 表示背叛选择。)

囚徒困境这一名称来自一个虚构的情形,其中两名犯人共同犯下重罪,但是当局只有他们犯有轻罪的证据。当局为每个囚犯提供了一项交易:检举另一位以获取从轻处罚。如果只有一个囚犯检举,那么他将得到非常轻的刑罚,而另一位则判重刑。如果两个人都检举,他们都将获得中度重判。如果没有人检举,每个人都将获得适度轻判。

在研究中,囚徒困境的收益通常是正的而不是负的。表 1 给出了标准的收益。

重复的囚徒困境博弈有过反复试验,两名玩家在每次试验结束时获悉另一位的做法。在一个重复的囚徒困境中,玩家可以策略性地运用自己的行为来影响另一位玩家的行为。对于自然资源的社会竞争也被认为是一种重复的囚徒困境,试验中产生最高收益的策略应该是受自然选择偏爱的。于是,阿克塞尔罗德和汉密尔顿(Axelrod & Hamilton,1981)邀请了经济学、社会学、政治学和数学中的博弈理论家来提出策略,然后在重复的囚徒困境博弈中彼此竞争。在他们提议的 14 个常用的相当复杂的策略中,最好的策略被证明是以牙还牙。这个策略很简单;它包括在第一次试验中合作,然后采取另一位玩家在前一个试验中的行动方式。以牙还牙的成功导致一些人认为,报复的倾向是我们进化过程中遗产的一部分。

单次囚徒困境博弈试验是用来研究社会动机,而不是策略。博弈理论假定每个人的唯一动机是个人利益最大化,预测人们在一次囚徒困境试验中普遍是背叛的。然而,实际上当人们被放置在一个囚徒困境试验中,有 1/3 到 1/2 的合作(Poundstone,1992)。这一发现对个人利益最大化是人际行为唯一动机的假设产生质疑。

——C. Daniel Batson

(叶鹏飞译 高勇校)

参考文献

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*, 1390-1396.

Poundstone, W. (1992). *Prisoner's dilemma.* New York: Doubleday.

Rapoport, A., & Chammah, A. M. (1965). *Prisoner's dilemma.* Ann Arbor: University of Michigan Press.

隐私(Privacy)

参见伦理准则【Ethical Codes】、伦理原则【Ethical Principles】。

隐私和保密(Privacy and Confidentiality)

A孤身一人,是位孤独的老人,并且健康状况每况愈下。她很高兴有一位礼貌地说普通话的研究者来访,询问她的健康和住房状况。A准备了茶水和饼干,诉说她的健康问题,直到她想到城市财政问题可能会通过减少健康恶化的贫困老人的福利来解决。为了防止这种情况发生在她的身上,她解释说她的(不存在)孩子负责她的医疗卫生需求,并且她可以根据自己的需要自由选择是否与他们一起生活。然而,研究者只不过是一个想要了解A所在社区需求的博士生。

A采取什么方式来控制访谈者与她的接触?什么造成了A最初接受访问后来又缩短接触时间?A关于保密的认知是什么?知情同意的内涵又是什么?

隐私和保密有多种不同的界定。这里的隐私是指个人及其对他人接近自己的方式的控制。保密是指个人自身的资料以及谁可以利用这些信息的协议。这些定义考虑了隐私的主观性质。有人乐意将自己的行为公开展示给他人,其他人则希望对他人隐藏行为,关于这些行为的资料也是如此。

有效的和合乎道德的研究需要对影响研究参与者判断隐私的因素具有敏感性。不能识别和尊重这些因素,可能会导致拒绝参与研究或影响回答的真实性,而很多敏感的或个人的问题需要进行研究。研究者如何学会充满尊敬地调查这些问题?关于人的研究的规则似乎是普适性的,很少提供明确的或程序上的指引。但是,这些规则引导研究者去寻找解决隐私问题的方法。根据研究的背景,涉及隐私的困境可能是具体的,需要特别的洞察力来了解特定研究总体的隐私利益,并且需要找到尊重这些利益的方式。下文旨在指引对这种洞察力的追求。

隐私

隐私的意义取决于个体的文化和个人环境、研究的性质以及研究所处的社会和政治环境。劳弗和沃尔夫(Laufer & Wolfe, 1977)提出一个关于个人隐私的理论,辨识出个体用来精心管理其隐私的各种因素,包括多重文化的、发展的和情景的因素。该理论认识到人们渴望控制在何时何地向他人提供何种信息,渴望控制他人提供给自己何种信息和经验。这一理论体现了以下四种隐私维度。

1.自我维度

自我维度是指自主发展和个人尊严。年幼的儿童不喜欢独自一人或者与他们的父母分开。在研究场景中,他们将父母看成是受欢迎的存在,而不是对其隐私的侵犯。到了童年中期,儿童寻求独处的时间来建立自我意识和培育新想法,这成为自尊、人格力量和尊严的基础;他们要求隐私,有时表现为秘密的藏身处和贴在卧室门上的“禁止入内”。青少年具有强烈的私人、自我和自恋意识,他们制造一种与父母相分离的身份。当有外人尤其是他们的父母在场时,访谈涉及个人私事会让他们感到尴尬。成年人也保持对独处时间的需求,而且会在巧妙地保护隐私方面变得更熟练(在前面的故事中,当A决定那样做是她的最大利益时,就转向了礼貌的谎言)。

2.环境维度

环境维度包括社会物质、文化和生命周期维度。社会物质要素是提供隐私的物质

和技术要素(如私人空间、电话)。文化要素包括实现隐私的标准(如善意谎言的文化可接受性)。生命周期要素随着年龄、职业、可利用的技术和不断发展的经济和社会文化模式而变化。

3.人际维度

人际维度是指互动和信息是如何管理的。社会场景及其物理特征为管理社会互动提供了选择;物理和社会边界可以用来控制人们与他人的接近。

4.控制选择维度

控制选择维度产生于其他三个维度。人们逐步学会运用个人、文化和物质资源来管理他们的隐私。在控制/选择过程发展的早期可能会威胁人们隐私的事件,后来就容易控制并且不会对隐私造成威胁。

用上述维度来指引调查,与熟悉研究总体的人际网络进行讨论可以提供有用的信息。这种网络可能包括相关的当地研究人员、教育工作者和社会服务人员(如社会工作者、农场代理人、公共卫生护士和其他专业人员)。焦点小组和其他形式的团体会议是非常有效的方法,可以用来学习如何使一个特定社区形成对研究的正确认知,如何使那些本来为他们所反感的研究能够为大家所接受,以及如何征集和通知预期的被访者(参见 Melton, Levine, Koocher, Rosenthal, & Thompson, 1988)。与正式的或临时的守门人及利益相关者交谈,志愿参与相关的社区服务活动,或者运用民族志的方法,也都能为研究总体的隐私关注提供有益的洞察力。就像上面列举的 A 的故事,理解和尊重被访对象的隐私利益有助于做好研究项目,因为人们有大量的技巧来秘密地保护他们的隐私,进而干扰研究者的目标。

保密

保密是隐私概念的扩展。它是指可辨认的资料(如被访者的录像带、实名的调查数据、包含足够的描述可以辨认出被访者的资料)和如何利用这些资料的协议。研究计划应考虑保密的需要以及哪些人有权获得这些资料。这应包括方法计划、机构审查委员会(Institutional Review Board, IRB)协议、知情同意、资料管理、研究助理培训和所有后续资料使用者的合约协议。可能的话,保密问题应通过以匿名方式收集资料,换言之,绝不记录下被访者的身份。身份标识包括可以把资料和一个特定的人联系起来的所有信息,诸如地址、车牌号码或社会保险号码。当个人资料汇集以后,被访者可能对接下来的防止其信息泄露的步骤尤其感兴趣。

然而,保密承诺不一定能增强一些人参加研究的意愿,对一些不敏感的研究作出广泛的保密承诺实际上反而增加了人们的怀疑,减少了参与的意愿(Singer, von Thurn, & Miller, 1995)。许多被调查者,尤其是少数民族,并不相信对保密的承诺(Turner, 1982)。被访者不信任问题的创造性解决方案,在于与研究相关的有益服务中与可信任的社区领导人进行协商与合作。

保密应当涉及谁或什么内容?关于人的研究的规则将人类受访者(human subjects)界定为独特的可辨认的(非匿名的)生命个体。然而,研究者因为诸多目标,也会考虑整个社区、组织甚至是死者及其亲属的隐私利益(参见 Andrews & Nelkin, 1998)。职业礼貌和保持与研究守门人之间的友善关系,也许和保护个体被访者一样,对科学和个体研究者非常重要。

保密的问题有很多意想不到的后果,就

像下面的小故事所说明的：为了回应资助机构的要求，弗雷谢尔教授在他的计划中申明，一旦他的研究成果出版，就将自己的资料向其他科学家开放。这样，其他人可以在他的资料的基础上进行研究，也许能把它们与来自相同样本的相关资料连接起来，或者采用不同的方法对他的原始资料进行再分析。

"不要那么快，弗雷谢尔教授！"他的IRB说，"你的项目说明中有一个麻烦的矛盾之处。在你的知情同意声明中，你承诺了保密，但是如果其他科学家要连接你的数据，你就必须附上唯一标识符。数据共享的确提高了效率并且减少了研究成本，但这并不能证明违反保密原则是正当的。"

弗雷谢尔教授对他的协议进行了以下修订：将使用更严格的保密标准来识别共享的数据，采用由美国人口普查局和其他联邦机构提出的复杂方法，允许将资料公布给学术研究者和一般公众以便进行更深入的分析。可以共享的数据资料有两类：列表数据（分组资料的列表）和微观数据（个体被访者的数据文件，显示每个入选变量的值）。同时，也有确保保密的各种方法：(a)在资料公开或存档之前进行去识别化（受限资料）；(b)对随后资料使用的人、时间和地点进行限制（限制访问）；(c)或者上述两种方法的结合。预计微观数据文件会在仔细审核之后向其他合格的科学家开放，使用保密和数据存取委员会（Confidentiality and Data Access Committee，CDAC）制订的建议公开数据的可能披露清单（Checklist on Disclosure Potential of Proposed Data Releases）。使用统计方法联邦委员会（Federal Committee on Statistical Methodology，FCSM）"入门书"（Chapter 2，FCSM，1994）中阐述的适当技术，执行对数据资料的限制。如果合格的科学家因为研究需要提出使用唯一标识符，将专门雇用统计人员根据"限制访问"的安排在这所大学中进行必要的分析，他们熟悉这些资料并且为二手用户组织和进行这些分析。

问题

为了防止被访者的信息被草率地泄露，要求对全体研究人员进行培训，并做好确保原始资料安全的计划。唯一标识符要尽快从资料中移走，而且必要的话，用代码来替代；代码的关键连接名称应保存在其他安全地点。

除了研究者自身的疏忽而致信息泄露以外，还存在各种各样威胁保密的可能性：

(1)基于资料中包含的人口学信息推断身份；

(2)个体被访者独特特征的识别（如来自录像带、定性的描述、个案研究）；

(3)黑客入侵存储有唯一标识符的计算机数据文件；

(4)资料的传唤；

(5)调查者法律强制的报告（如关于虐待老人或儿童）；

(6)相关数据集的连接（从不同的来源或纵贯性的序列）；

(7)研究资助者对数据的审核。

研究者有责任识别对他们计划收集的数据的保密可能存在的威胁，使用适当的方法确保保密，并向潜在的被访者告知如何实现保密的信息或者是否还存在影响保密的风险。研究者不能承诺他们不能保证的保密。

解决方法

有文献描述了确保保密的方式。下文概述确保保密的主要途径。每种途径都伴

有两个信息来源:(a)关键的参考书目;(b)在线搜索引擎上采用的关键词,诸如谷歌。随着新的保密问题和解决方法的出现,网络已变成一种快速发展的来源。

1.并入大类和掺杂误差

当某个人了解一个个体独特的识别性的人口学特征(如年龄、收入、医疗状况、邮政编码),希望探知该个体的其他信息,那么推断出身份是可能的。这可以通过并入大类来使独特个体的资料变得模糊来进行防止(例如,一个亿万富翁可以被包含在标签为"年收入超过 $100 000"的分类中),或者资料可以被"掺杂误差"以至于没有个案是必然适当的。(Johnson, 1982; 保密原始资料)

2.个体被访者的独特特征或个案描述和录像带中的场景

这些可以通过各种方法进行伪装,但是消除所有可能的识别性特征,会破坏资料的科学价值。第二项预防措施是假设身份能够被猜测到,因此避免使用贬义词或不必要的个人语言。(Johnson, 1982; 保密定性资料)

3.资料的传唤

资料有可能被当作是法律诉讼的材料。在美国,如果国家卫生机构的研究者人员事先收集了可能被传唤的资料,他们可以用保密证书阻止传唤。很明显,被访者犯罪活动的资料需要一个保密证书。与未来集体诉讼案、儿童监护纠纷或其他公民法律行为有关的资料是否需要保密证书则不易判断;这些对保密的威胁不可能预料到。保护传唤资料的一个更加安全的方式,是尽可能快地移除唯一标识符。(National Institutes of Health, 2003; 保密,资料传唤)

4.资料和唯一标识符之间的连接

如果可能,匿名是最好的做法。然而,与标识符之间的连接可能是需要的,例如需要对抽样效度进行后续检验或开展纵贯性调查。临时性地识别回答者,代理人、化名或文件链接系统的应用,这些都能够既不泄露被访者身份又实现资料的连接。(Campbell, Boruch, Schwartz, & Steinberg, 1977; 保密文件连接)

5.资料审核

研究资助者的资料审核,要求资助者能够再次接触被访者,以确保数据确实是从相应的被访者收集起来的。为满足这一要求同时又要保护被访者的隐私利益,研究者(或研究机构)可能要围绕如何确保保密与这类二级用户签订合约协议。研究者应该在知情同意中公开这些协议,这样潜在的被访者就能了解研究者在项目之后将会如何处理这些资料。(Boruch & Cecil, 1979; 保密资料共享)

6.法律强制报告的要求

法律强制报告有关虐待儿童的事件,但是关于什么构成虐待和谁应当强制报告,这在美国各个州和县是有差异的。这些要求通常针对的是能够提供帮助的专业人士(治疗师、医师、教师);很多研究者恰好也正是这类人。此外,在九个州——佛罗里达州、印第安纳州、肯塔基州、明尼苏达州、内布拉斯加州、新罕布什尔州、新泽西州、新墨西哥州和北卡罗来纳州——任何人只要有理由怀疑虐待儿童就可以举报。虐待儿童可能包括各种不同的行为,例如疏忽、身体伤害、心理伤害、性虐待和儿童色情作品。研究的知情同意中可能会透露可报告的事件,所以应当陈述保密的承诺,除非研究者对儿童受到虐待有合理怀疑。对儿童受到虐待有合

理怀疑的研究者,在向地方当局报告事件之前,应当咨询对报告要求有更深理解的专业人士(如护士、社会工作者、医师)。(Kalichman, 1999; 报告虐待儿童)

关于预防和报告虐待老人的法律正在迅速完善中。大多数虐待老人的行为都是家庭成员犯下的。它包括拍打、伤害、性虐待或行为控制;为了其他人的利益而未经他们同意的经济剥削;没有提供必要的物品和服务来避免老人的身体伤害、精神创伤或精神疾病。(Bonnie & Wallace, 2002; 报告虐待老人)

7.计算机文件的保护

研究者通常将电子资料存储在他们机构的服务器上。这些数据容易受到黑客、冒充者、局域网用户和特洛伊木马的攻击。将带有唯一标识符的数据存储在未授权用户可以获得访问权限的地方是不明智的。对这些风险和相关防护措施的了解也在迅速发展;当前最好的信息来源是网络。(Skoudis, 2002; 计算机文件黑客;保密)

8.在线研究

调查、焦点小组、人格研究、实验、民族志和其他各种社会和行为研究正越来越多地在网上进行,这使研究者能够迅速地接触遥远的被访者,并且花费不多、可以连续进行以及不需要额外的研究人员。保密议题所伴随的问题和解决方案会随着新技术的发展而迅速改变。保护通信和个人身份的新技术以及技术高明的隐私保护团队开始出现。然而,数据资料保护的政策和方法发展不够迅速,难以应对新出现的问题。目前已出现全球范围内的数字通信网络;黑客们利用一台笔记本电脑和网络技术就可下载到存储在世界各地任何服务器的任何电子数据。(Nosek, Banaji, & Greenwald, 2002; 网络基础的研究)

启示

社会科学内部存在一种固有的认识,即被访者的隐私应当会得到尊重,应当适当限制获取可识别资料的途径。但是什么是被访者的隐私利益?什么样的限制才是适当的?如果存在特殊的隐私利益,那么可能需要特殊预防措施(如保密证书、错误移植)以及应当与被访者和守门人商谈。当被访者与研究者处境不同时,这个问题会更复杂,因为很难洞察被访者的处境及其隐私利益。例如,被访者会相信研究者的保密承诺吗?有没有"好事者"会去推断识别被访者或者谁将会传唤数据?澄清此类问题是很困难的,并且很难用标准的语言或在条款中明确地解决。这些议题需要研究者熟知其研究对象的文化和情境。

——Joan E. Sieber

(叶鹏飞译　高勇校)

参考文献

Andrews, L., & Nelkin, D. (1998). Do the dead have interests? Policy issues of research after life. *American Journal of Law & Medicine*, *24*, 261.

Bonnie, R. J., & Wallace, R. B. (2002). *Elder mistreatment: Abuse, neglect, and exploitation in an aging America*. Washington, DC: Joseph Henry Press.

Boruch, R. F., & Cecil, J. S. (1979). *Assuring the confidentiality of social research data*. Philadelphia: University of Pennsylvania Press.

Campbell, D. T., Boruch, R. F., Schwartz, R. D., & Steinberg, J. (1977). Confidentiality-preserving modes of access to files and to interfile exchange for useful statistical analysis. *Evaluation Quarterly*, *1*, 269-300.

Confidentiality and Data Access Committee. (1999).

Checklist on disclosure potential of proposed data releases. Washington, DC: Office of Management and Budget, Office of Information and Regulatory Affairs, Statistical Policy Office. Retrieved from www. fcsm. gov/committees/cdac/ checklist 799.doc.

Federal Committee on Statistical Methodology. (1994). *Report on statistical disclosure limitation methodology* (Statistical Policy Working Paper #22). Prepared by the Subcommittee on Disclosure Limitation Methodology. Washington, DC: Office of Management and Budget, Office of Information and Regulatory Affairs, Statistical Policy Office. Retrieved from www. fcsm. gov/working-papers/ wp22.html.

Johnson, C. (1982). Risks in the publication of fieldwork. In J. E. Sieber (Ed.), *The ethics of social research: Fieldwork, regulation and publication* (pp. 71-92). New York: Springer-Verlag.

Kalichman, S. C. (1999). *Mandated reporting of suspected child abuse: Ethics, law and policy*. Washington, DC: American Psychological Association.

Laufer, R. S., & Wolfe, M. (1977). Privacy as a concept and a social issue: A multidimensional developmental theory. *Journal of Social Issues*, *33*, 44-87.

Melton, G. B., Levine, R. J., Koocher, G. P., Rosenthal, R., & Thompson, W. (1988). Community consultation in socially sensitive research. *American Psychologist*, *43*, 573-581.

National Institutes of Health. (2003, July 22). *Certificates of confidentiality*. Retrieved from http://grants1. nih. gov/grants/ policy/coc/ index.htm.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). E-research: Ethics, security, design, and control in psychological research on the Internet. *Journal of Social Issues*, *58*, 161-176.

Singer, E., von Thurn, D., & Miller, E. (1995). Confidentiality assurances and response. *Public Opinion Quarterly*, *59*, 66-77.

Skoudis, E. (2002). *Counter Hack: A step-by-step guide to computer attacks and effective defenses*. Upper Saddle, NJ: Prentice Hall.

Turner, A. C. (1982). What subjects of survey research believe about confidentiality. In J. E. Sieber (Ed.), *The ethics of social research: Surveys and experiments* (pp. 151-165). New York: Springer-Verlag.

概率的(Probabilistic)

如果一个模型或命题是随机的【Stochastic】(即包含不确定性)并运用了概率【Probability】的概念,那么它就是概率的。当研究者没有足够的信息运用确定性的【Deterministic】方法,而可获得的信息最便于根据概率进行描述时,在科学研究中通常使用概率模型和命题。一个社会现象的概率模型是这样一组假设,根据人际关系、关联、因果关系或其他具有自身概率的概念来描述各个实体的可能行为。概率命题通常基于概率模型,其目的不在于证明一个特定事实是真实的或者这个事实是所讨论行为或发展的结果,而对特定后果和结果的概率进行评估。

应用在概率模型或命题中的概率可能是主观考虑、既有知识或经验的结果,或者模型分析的结果,以及通常用来评估概率是否有现实依据的那些命题的结果。在这种

情况下,概率命题被用来论证概率命题。

在社会科学中,概率模型和命题的应用有两种主要情况。其一,支配两个特定变量之间关系的确定规则未被精确了解。例如,在大多数社会中,父亲受过高水平教育的孩子相对于父亲受教育水平较低的孩子,有更好的机会接受更高水平的教育。基于教育流动数据,一个给定父亲教育水平的孩子获得基础教育、中等教育或者高水平教育的条件概率可以被估计出来。根据其他潜在地影响教育获得的因素,诸如母亲的教育水平、种族或家庭宗教背景、家庭生活居住地的类型等,这些概率可以得到进一步的修正。通过考虑这几个因素的影响,概率可以得到完善,产生一个对真实情况的良好近似。这是一个概率模型,因为并不是所有因素都被考虑进去,并且其影响也无法精确了解;更确切地说,它们是用概率来描述的。从概率模型作出的预测可以与现实进行比较,以获得对模型中的概率更好的估计并且找出是否需要深入考虑的因素。

概率命题和模型应用的另一个重要情况是,确切的因素及其影响只是次要的关注,研究者更关心基于样本【Sample】数据来预测【Predicting】整个总体【Population】的行为。一个典型的例子是,运用民意调查的数据预测选举结果。例如,如果总体中25%的人愿意投票选择某一候选人,那么在样本容量为2 500时,民意测验专家有0.95的概率得到一个23%~27%的结果。基于这一事实,如果民意调查结果是该候选人有26%的欢迎度,那么民意测验专家能够得出真实值为95%的置信区间【Confidence Interval】(24%,28%)。

——Tamás Rudas

(叶鹏飞译 高勇校)

参考文献

Bennett, D. J. (1998). *Randomness.* Cambridge, MA: Harvard University Press.

概率性的哥特曼度量(Probabilistic Guttman Scaling)

概率性的哥特曼度量是经典的确定性的哥特曼度量【Guttman Scaling】方法的扩展。这种方法中存在随机性,它有很多类型,其中最突出和最早的是由普罗克特(Proctor,1970)提出的方法。

——Alan Bryman

(叶鹏飞译 高勇校)

* 也可参见哥特曼度量【Guttman Scaling】。

参考文献

Proctor, C.H. (1970). Aprobabilistic formulation and statistical analysis of Guttman scaling. *Psychometrika*, *35*, 73-78.

概率(Probability)

现代概率论是数学领域最广泛的应用之一。由于抽样【Sampling】,对实际过程或观察现象的支配规则理解不足,或其他原因,任何一个经验调查都会涉及不确定性,

因而可能应用到概率的概念。一般而言,概率被认为是特定观察事件发生的可能性大小的数字表示。

历史

概率的概念发端于对各种游戏特性的调查,比如掷骰子,但除了搞清楚如何获胜这一非常实用的愿望以外,其中也蕴含着深刻的哲学思想。首先,关于赌博中特定结果出现的机会问题的计算,引导出有关概率规则的计算,但概率本身的概念并没有界定。很长一段时间之后,人们才意识到概率背景或框架的存在并不是不言自明的,“概率是什么?”变成一个核心问题。在19世纪晚期,哲学家和数学家广泛接受的是,用来定义一个精确理论(不仅是对概率而且针对很多其他目前我们作为科学思维基础的概念)基本原理的正确方法是公理化方法。公理化方法的目的不是精确地定义什么是概率;确切地说,它先对其属性做出假定并从中推导出整个理论。柯尔莫哥洛夫在20世纪初提出了概率的公理化理论。虽然关于概率本质的哲学思考持续到现在,但柯尔莫哥洛夫的理论一直能够涵盖最新的发展,并且这一理论在统计学中引导出广泛的应用成果。

例子和公理

通常,若干结构拥有在公理体系中假定的属性,它们可以作为概念的模型,但概念本身并非来自于此,而只是由这些构念【Construct】来说明。对于概率,一个简单的例子就是掷骰子,所有结果出现的可能性是相等的。每个结果的概率都等于1/6,结果小于4(即1,2,3)的概率是1/2,而出现偶数结果的概率(即2,4,6)的概率也是1/2。要求结果小于4且为偶数的概率等于1/6,那么这个数只能是2。然而,有些事件可能出现但是不可能同时发生。例如,要求小于4且大于5的值是不可能的。不可能发生的事件被称为不可能事件,记为0。

另一个例子是所谓的几何概率,它是通过打靶的方式,射中任一点都具有相同的概率而得到的模型。这里,射中右半部分的概率是1/2,正如射中左半部分或上半部分的概率一样。要求同时射中右半部分和上半部分的概率是1/4。

在一个精确理论中,与一次实验的观察相关的事件具有概率。要定义概率的性质,就必须定义这些事件的特定运算。如果同时出现A和B,A和B两个事件的乘积记作AB。例如,在掷一个骰子时,一个事件的取值可能小于4,另一个可能是偶数值。那么,两者的乘积就是既小于4又是偶数值,也就是2。或者,一个事件可能是射中靶子的右半部分,另一个可能射中它的上半部分,那么乘积是射中右上半部分1/4的范围。如果两个事件不能同时发生,它们的乘积就是不可能事件。另一个运算是两个事件之和,记为$A+B$。只有当两个事件中至少有一个发生时,这个事件的和才会出现。小于4的事件与具有偶数值的事件之和,是来自于1,2,3,4,6中的任何事件。靶子的右半部分与它的上半部分之和是靶子的3/4,可通过排除左下方1/4获得。

我们还可以考虑事件的更复杂运算。例如,将靶子的右上角表示为A_1,直接在其之下的靶子的1/8表示为A_2,A_2下面的1/16表示为A_3,A_3下面的1/32表示为A_4,以此类推。这里会有无限多的事件,并且直观上看它们也有一个和,即靶子的左半部分。如图1所示。这些无限多的事件之和可以用$A_1+A_2+A_3+A_4+\cdots$表示。概率论也同样适用于这些更一般的运算。

现在我们来用公式表示概率的公理。概率是分配给与一个实验相关联的事件的一个数字(概率)系统。事件的概率分配方式为:

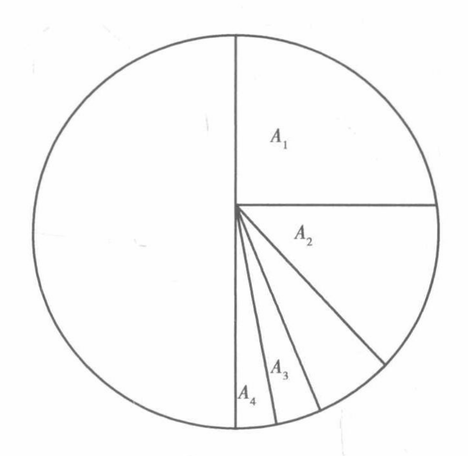


图 1 一个靶子上呈现的无限多事件之和

①一个事件的概率总是在 0 和 1 之间：

$$0 \leqslant P(A) \leqslant 1$$

②如果一个事件始终（肯定地）发生，它的概率是 1：

$$P(S)=1$$

③如果两个事件不可能同时发生，那么它们和的概率等于它们的概率之和：

$$P(A+B)=P(A)+P(B)\text{，如果 } AB=\varnothing$$

最后一个性质可以扩展到无限多事件之和，正如在射中靶子特定部分的例子中一样。公理 3 的一般式假定如下：

④如果一个事件序列 $A_1, A_2, A_3, \cdots$ 中，任何两个事件都不可能同时发生，那么它们和的概率就是它们的概率之和。

$$P(A_1+A_2+A_3+\cdots) = P(A_1)+P(A_2)+P(A_3)+\cdots$$

如果 $A_iA_j=\varnothing, i \neq j$。

独立性

概率的所有性质都可以从公理 1 到公理 3（或 1～3′）中推导出，而且概率论中更深入的概念正是在这个框架中进行定义的。概率论中一个常用的概念是独立性。如果

$$P(AB)=P(A)P(B)$$

那么事件 A 和 B 是独立的。

需要注意的是，此公式不是在独立性为真时用来计算联合概率的；它是独立性的定义。在投骰子的例子中，出现偶数值的事件与数值小于 5 的事件是独立的，因为前者（2,4,6）的概率为 1/2，后者（1,2,3,4）的概率为 2/3，而交集（2,4）的概率为1/3，即为两者的乘积。另一方面，出现偶数的事件与小于 4 的事件并不是独立的，因为前者的概率为 1/2，后者的概率也为 1/2，而它们乘积的概率是 1/6。这一事实的解释可能是，前三个数字中偶数出现的可能性，要比前六个数字中偶数出现的可能性低。确实，在前六个数字中有三个是偶数，但在前三个数字中只有一个偶数。

条件概率

事实上，后者的解释指向另一个与概率相关的有用概念，称为条件概率。条件概率表示在知道一个事件发生后，另一个事件变化的概率。例如，在掷骰子时，出现偶数的概率是 1/2，因为所有可能结果的一半是偶数。然而，如果一个人知道小于 4 的事件已经发生，那么知道这一点后，发生偶数值的事件概率是不同的，这个新的概率被称为给定数字小于 4 的情况下，出现一个偶数的条件概率。为便于理解，来看一个实验的可能结果的列表：

可能的结果：1,2,3,4,5,6
其中出现的偶数：2,4,6
出现一个偶数的概率：1/2
数字小于 4 的可能结果：1,2,3
其中出现的偶数：2
数字小于 4 后出现偶数的条件概率：

1/3。

给出 B 后 A 的条件概率记为 $P(A|B)$，可以精确地定义为：

$$P(A|B)=\frac{P(AB)}{P(B)}$$

即事件乘积的概率除以作为条件的事件概率。在上面的示例中，我们已经看到，出现的数字是偶数且小于 4 的概率是 1/6，出现数字小于 4 的概率是 1/2，它们的比率为 1/3。

从定义直接推断，如果 A 与 B 是独立的，那么 $P(A|B)=P(A)$；也就是说，如果它们是独立的，那么以 B 的发生作为条件，不会改变 A 发生的概率。事实上，如果 A 和 B 是独立的，那么

$$\begin{aligned}P(A|B)&=\frac{P(AB)}{P(B)}\\&=\frac{P(A)P(B)}{P(B)}=P(A)\end{aligned}$$

其中，第一个等式是条件概率的定义，第二个等式是以独立性的定义为基础。因此，事件的独立性意味着彼此之间没有相关性。如果 A 与 B 是独立的，无论 B 是否发生，A 的概率都是相同的。

在打靶的例子中，通过类似掷骰子的例子中提出的论证方法很容易看出，射中右半部分和下半部分的事件是彼此独立的。这在很大程度上依赖于这样一个事实，这两个例子是相似的，其中所有结果出现的可能性都是一样的。在投掷一个骰子的案例中，结果是数字 1，2，3，4，5，6；在打靶的例子中，结果是靶子上的落点。尽管人们倾向于接受这两个例子中的所有结果都具有同样的可能性，但是这一假设对于掷骰子中的 6 个结果来说更为直观，而对于打靶中无限多的可能结果来说不太明显。这显示了基础性数学理论的力量，这种假设可以精确地纳入概率模型。在社会科学中，具有无限多实验结果的模型是经常使用的；例如，基于正态分布【Normal Distribution】的近似法假定一个变量可以取任何数字作为它的值，并且存在无限多的不同数字，正如靶子上无限多的点。在这种情况下，人们采用所谓的密度函数来描述概率行为。

应用

上述的简单例子并不意味着公式化概率模型和应用概率模型总是那么简单直接。有时，即使是相对简单的模型分析起来也相当复杂，但是，更重要的是，人们很容易在基本概念层面上犯错。下面的例子说明了一些这样的危险。一个公司 CEO 面试三个申请秘书职位的应聘者。应聘者为 A，B 和 C，他们只知道有三个应聘者，并没有彼此资历的相关信息，可以假设他们都有同等的机会获得这份工作。因此，他们都认为每个人被录用的概率是 1/3。面试之后，CEO 作出了决定，但这个决定还没有告知应聘者。应聘者 A 打电话跟 CEO 说：“由于只有三个面试者，而其中一人将获得职位，可以肯定的是 B 和 C 之中至少将有一人落选。”CEO 没有给出 A 获得职位可能性的信息，而是说出 B 和 C 中哪一个没有获得职位。CEO 告诉 A，C 肯定不会得到职位。通话结束后，A 现在认为，C 已经退出竞争，由于只剩下两个候选人 A 和 B，每人就有 1/2 的概率获得这个职位，也就是说，概率从1/3变为 1/2。关于 A 的机会，CEO 是否通过指明一个不会得到这份工作的人（如果有这样的人存在）而给出了信息？A 得到这份工作的正确概率是 1/3 还是 1/2？这种情况是有点矛盾，因为人们可以同意也可以反对这两个问题的两个可能的回答。这里重要的一

点是，假设有三个应聘者（概率是 1/3）或假定有两个应聘者（概率是 1/2）是两个不同的概率模型，因此不能直接进行比较。更确切地说，追问 1/3 或 1/2 哪个才是正确的概率，只有当他们在同一个模型中计算时才有意义（在这种情况下才有一个明确的答案）。在第一个模型（三个应聘者）中，指明那个没有得到工作职位的应聘者确实没有多大信息量。但是，当 CEO 指出那个没有得到工作职位的应聘者，A 就可以根据这一信息，定义一个新的概率模型，并且在新的模型（两个应聘者）中，正确的概率是 1/2。要使基于概率模型的推断（这也适用于任何类型的统计推论【Statistical Inference】的）是适当的，前提是除了模型本身的正确分析外，还要以模型恰当地描述了相关事实为基础。结果将在很大程度上取决于模型的选择。

概率论在社会科学领域有许多应用。它是随机抽样【Random Sampling】的基础，在随机抽样中，感兴趣的总体中的单元（通常是人）是以事先指定的概率被选取到样本中。最简单的情况是简单随机抽样【Simple Random Sampling】，其中每个人都有相同的机会被选进样本，并且每个人被选择的过程是相互独立的。在这种情况下，概率模型的所有可能结果都具有相等概率，模型可能是适当的，但在实践中运用更复杂的抽样方案往往需要其他模型。概率方法也被用来对测量误差的影响进行建模，更普遍的情况是用来对没有测量到的或未知因素的影响进行建模。

——Tamás Rudas

（叶鹏飞译　高勇校）

参考文献

Bennett, D. J. (1998). *Randomness*. Cambridge, MA: Harvard University Press.

Feller, W. (1968). *An introduction to probability theory and its applications* (3rd ed.). New York: John Wiley.

Wonnacott, T. H., & Wonnacott, R. J. (1990). *Introductory statistics for business and economics* (4th ed.). San Francisco: Jossey-Bass.

概率密度函数（Probability Density Function）

连续变量【Continuous Variable】的概率行为最好是用它的概率密度函数来描述。尽管对于离散变量【Discrete Variable】，其对应于不同值的概率可以分别指定，但这对于连续变量来说是不可能的。连续变量所有可能值对应的概率【Probability】是 0，因此，指定单个概率没有什么意义。有用的以及事实上能够完整刻画一个连续变量行为的，是在一个任意区间内取值的概率。一个观察值出现在一个区间内的概率可以通过概率密度函数在这个区间上的积分确定，也就是概率密度函数在这段区间上的面积。图 1 可以说明这个过程。阴影部分面积就是一个观察值落在区间（a, b）上的概率。其他区间的概率同样也可以确定。这个概率取决于区间的长度［更长的区间（c, d）有更大的概率］和区间的位置。例如，区间（e, f）和区间（a, b）有相同的长度，但是观察值落在区间（e, f）上的可能性要比落在区间（a, b）上的可能性大。很容易看出，通过这种方法确定的区间概率具有概率的性质。例如，如果一个区间是由两个非重叠区间合并起来的，那么与它相关的面积是这两个区间的相关面积之和。

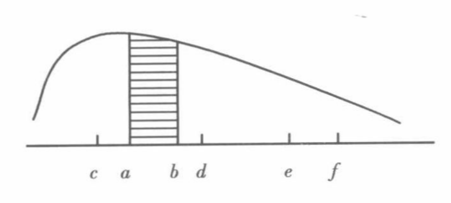


图 1 概率密度函数

通过计算相关的面积,密度函数也可以用来确定一个变量大于或小于某个特定值的概率。密度函数除了其他性质以外,必须是非负的,而且函数曲线下的总面积等于 1。

密度函数可以被认为是直方图【Histogram】的理论化扩展,通过将直方图的区间变得越来越窄而得到。与此同时,区间的数量增加,直方图的边界变得越来越细。

密度函数的名字来源于一个事实,即某个区间上的概率确定了这个区间上预期观察值的密度。当观察一个变量时,在某个区间上预期观察值的数量与该区间的概率是成比例的:在一个概率较高的区间上,观察值的出现要比在较低概率的区间上更密集。

变量某个可能取值的密度函数值被称为似然值(likelihood of the value)。例如,取值 a 的似然值要比 e 的似然值大。在日常用语中,概率(probability)和似然(likelihood)可以互换使用,但在概率和统计中,它们具有不同的含义。a 和 e 这两个值的概率均为零,但它们具有不同的似然值。这种差异的一个准确解释是,在 a 附近的值要比在 e 附近的值更有可能发生。还有一个事实是,以 a 为中心的一段小区间要比以 e 为中心的同等长度的区间有更大的概率。

——Tamás Rudas

(叶鹏飞译 高勇校)

参考文献

Chatfield, C. (1983). *Statistics for technology* (3rd ed.). London: Chapman & Hall.

概率抽样(Probability Sampling)

参见随机抽样【Random Sampling】。

概率值(Probability Value)

概率值(或 p 值)或得到的概率水平测量的是数据能够提供多少证据用以拒绝统计假设【Hypothesis】。它表示如果假设成立,观察到这种极端数据的可能性有多大。p 值本质上就是能够观察到和实际数据一样极端以及比实际数据更为极端的情况的概率。假设检验通常是以 p 值为基础的。如果得到的概率水平小于一个指定的值——通常是 0.05 但有时是 0.01 或 0. 10,那么就拒绝原假设。大多数软件包通常都会输出 p 值,通过将这些 p 值与期望的显著性水平【Significance Level】相比较,可以进行显著性检验【Significance Testing】。

举一个简单的例子,考虑这样的二项检验【Binomial Test】,其中观察数量 $n=10$,原假设是成功的概率 $p=0.5$。观察值的一些属性中只有成功的次数与检验假设有关。表 1 列出了所有可能的结果及其对应的概

率。如果数据包含7次成功(3次失败),那么比当前结果更有可能的结果是成功6次、5次或4次。可能性与实际结果相等或比其更小的结果包括0,1,2,3,7,8,9或10次成功;它们的联合概率是0.347 5,这就是与数据和检验相关的p值。因此,实际数据(7次成功)不属于最小可能的5%观察值,不能拒绝原假设。另外,如果观察数据仅包含1次成功,那么可能性与实际结果相等或比其更小的结果包括0,1,9或10次成功,它们的联合概率是0.021 484。这个概率水平小于0.05,可以拒绝原假设,因为在假设成立的情况下,观察到这些数据的可能性极小。

表1 二项检验的可能结果($n=10$, $p=0.5$)

成功的次数	概 率
0	0.000 977
1	0.009 766
2	0.043 95
3	0.117 188
4	0.205 078
5	0.246 094
6	0.205 078
7	0.117 188
8	0.043 95
9	0.009 766
10	0.000 977

另一个例子是考虑2 × 2的列联表【Contingency Table】,原假设是变量独立【Independence】。数据如表2所示。在这个例子中,存在很多个可能的无关表格,样本的概率取决于哪一个表是真实的分布。在这种情况下,可以运用最大似然估计【Maximum Likelihood Estimation】来选择使实际观察值具有最大概率的分布,并依据这个分布确定p值。为了确定概率值,我们首先要计算卡方【Chi-Square】统计量。那些比实际数据具有更小似然值的样本,计算出来的统计量是一个更大的值。这些样本出现的概率(即p值)等于该统计量的取值大于它的实际值的概率。对于目前的数据,皮尔逊卡方统计量的值是0.793 651;统计量超过该值的概率是0.37,这就是该数据的p值。

表2 2×2表中假定的观察值

10	20
30	40

——Tamás Rudas

(叶鹏飞译 高勇校)

参考文献

Schervish, M. J. (1996). *P*-values: What they are and what are they not. *American Statistician*, *50*, 203-206.

追问(Probing)

追问,是一个在现场访谈期间用于获取额外信息的技术,即在问了一个开放式问题【Open-ended Question】之后进一步要求更多的解释或说明。它补充了访谈协议中的具体问题。这项技术可以使调查者获得研究项目重点关注的准确信息。

研究设计可采用不同类型的访谈形式,这些形式在结构程度上存在差异。在连续谱的一端,是高度结构化访谈【Structured Interview】表,访谈员向所有被调查者询问

相同的直接性问题;这些类型的访谈协议在设计上非常类似于邮寄【Mail】调查。一个著名的例子是全国选举调查(National Election Survey)。在另一端,是更为自由的焦点访谈形式,访谈员询问一些一般性的问题,然后让访谈从这些问题中发展出来;这样的访谈有一种自然对话的感觉。在结构化较低的访谈中,往往需要更多的追问。访谈员会询问一些一般性的问题,然后在谈话过程中追问被调查者以掌握感兴趣的信息。

与基本的访谈问题一样,追问绝不能引导或建议被访者的回答。追问应该简单地促使被访者去详细说明他或她的初始反应。在访谈中需要牢记,有时最适当的追问是一段时间的沉默;一个停顿往往能够促使被访者发现要点。被访者总是有区别的,有些人比其他人更为健谈,不需要太多的追问,其他人可能会相当沉默。例如,如果一个被调查者被问到他或她的组织是否会使用多种制度化的方法来公开组织信息,被访者简单地回答"哦,我们使用书本上记载的每一种方法,凡是你能想到的我们都尝试过",那么可以简单地追问,"比如?"或者"你能描述一下这个过程吗?"促使被访者作出更深入的解释。

——Christine Mahoney

(叶鹏飞译　高勇校)

参考文献

Berry, J. (2002). Validity and reliability issues in elite interviewing. *PS—Political Science and Politics*, *35*, 679-682.

Dexter, L. A. (1970). *Elite and specialized interviewing*. Evanston, IL: Northwest University Press.

Probit 分析(Probit Analysis)

Probit 分析起源于分析只有两种可能性的(二分的)响应的方法。在仅有两种可能性的响应(quantal responses)涉及的情况中,针对一个刺激只有一个可能的响应,有时被称为"有或无(all-or-nothing)"。Probit 分析早期应用于生物学实验。一个经典的例子是当某种昆虫暴露在某种杀虫剂时的死亡测定。虽然理论上来讲,杀虫剂的量越大,昆虫的健康水平越恶化,但是测量的局限性或只关注某一特定结果使得我们有必要使用二分变量【Dichotomous Variable】代表仅有两种可能性的响应。目前该分析在社会科学中经常应用于个体层面行为,诸如一名国会议员是否赞成或反对一项法案,或者一名法官什么时候会肯定或撤销下级法院的判决。

社会科学的分析工作经常使用线性回归分析的形式(如最小二乘【Ordinary Least Squares,OLS】)。回归分析的一个基本假设是因变量【Dependent Variable】是连续的,这样才允许建立线性模型。然而,许多研究问题涉及二分因变量的情况(如人类活动中的决策行为,最基本的例子是个人投票行为)。当因变量是二分变量时,由于违背了线性回归的假设,因而可能会导致推断【Inference】中出现严重错误,二分变量的分析一般需要非线性的估计过程,Probit 是一个常见的选择。

回归的概念方法是在给定自变量值的情况下,估计因变量的观察值和预测值之间的误差平方和最小的线性模型的回归系数。相比之下,Probit 的方法是在给定自

变量值的情况下,去估计能使观测到的因变量值似然性最大化的非线性模型的系数。因此,Probit 估计是通过最大似然估计【Maximum Likelihood Estimation】方法来获得的。

Probit 模型有四个基本假设。第一,误差项服从正态分布。在这一假设下,Probit 估计与 OLS 有一系列相似的特性(如无偏性、正态分布、最小方差)。第二,因变量取值为 1(给出的选择是 0 和 1)的概率是一个关于自变量的函数。第三,因变量的观察值是统计独立的。这排除了序列相关【Serial Correlation】。第四,在自变量中没有完美的线性相关。正如 OLS 一样,较强的线性相关可能会导致多重共线性【Multicollinearity】。

尽管 Probit 估计看上去类似于 OLS 回归估计,但它们的解释却不那么直接。Probit 估计是指对于某一特定观察,在给定自变量值的前提下,观察到因变量值为 0 或 1 的概率。每一自变量的系数估计值是因变量取值为 1 的概率的增加值(如果估计值为负值,则是减少),但是这一增加值是用累积正态概率分布中的标准离差(即 Z 值)形式表达的。因此,对于给定的观察,$y=1$ 的估计概率为:

$$P(y=1)=\Phi\left(\sum_{i=1}^{n}\hat{\beta}_0+\hat{\beta}_i x_i\right)$$

其中,$\hat{\beta}_0$ 为常数项的估计系数;$\hat{\beta}_i$ 为 n 个自变量(x_i)的估计系数;Φ 为累积正态概率分布。

表 1 显示了包括三个自变量的模型简单的 Probit 结果,每一个自变量都是二分(0/1)测量。对于一个特定观察,$y=1$ 的概率为:

$$P(y=1)$$

$$=\Phi[(2.31)\times常数+(-2.16)\times x_1+(-1.06)\times x_2+(-2.27)\times x_3]$$

表 1 Probit 估计的简单结果

因变量 y		标准误差
自变量	估计系数	
常数	2.31	0.56
x_1	−2.16	0.42
x_2	−1.06	0.45
x_3	−2.27	0.80
N	107	
$-2LLR$ 3 df	65.28	

如果所有三个自变量 x_i 的取值均为 1,Z 分数为−3.18。计算机程序会自动计算概率,或者可以查表确定,对于这个观察 $y=1$ 的概率仅为 0.000 7。如果全部三个 x_i 取值均为 0,Z 分数为 2.31,$y=1$ 的概率为0.989 6。

其中一个 x_i 值的改变所产生的边际效应,通常可以通过控制其他 x_i 的值不变,并基于相关变量来计算概率的变化而得到。其他 x_i 值的选择可以是它们的均值,尽管当变量实际上不能取这样的值或某些代表性的组合值时,这在理论上会存在问题。例如,如果我们想知道 x_1 的效应,那么可以设定 $x_2=0$ 和 $x_3=1$,那么当 $x_1=1$ 时,$y=1$ 的概率为 0.017 0,而当 $x_1=0$ 时,概率增加到 0.516 0。注意,当 Z 分数接近于 0 时,所观察的特定变量的表面效应将显得更大。例如,如果其他变量的 Z 分数等于 0,那么一个标准差的改变将使概率从 0.5 增加到 0.841 3。相比之下,如果开始时 Z 分数为 2.0,那么同样一个标准差的改变只会使概率从 0.977 2 增加到 0.998 7。

估计系数的显著性检验是通过 t 统计

量来实现的，t 统计量是把估计值除以标准误得到的。在前面的例子中，第一个变量的结果是 $-2.16/0.56=-5.14$，这在 $p<0.001$ 时是显著的。对数似然比可以用来检验模型的整体显著性。具体来说，在 $-2(\ln L_0-\ln L_1)$ 中，$\ln L_1$ 是全模型似然值的自然对数，$\ln L_0$ 为除常数项以外的全部变量为 0 时的值。对数似然比通常记为 $-2LLR$，是在除常数项以外的所有系数为 0 时，对假设进行检验的卡方值。自由度是自变量的个数。

Probit 结果的拟合优度量度【Goodness-of-fit Measures】有两种基本类型，第一种是伪 R^2，试图评估模型解释了多少方差。这类测量的一个问题是，它们没有考虑模型可以成功地将因变量的观察值正确地归为两个类别。当每种类别的观察数目悬殊时，问题会变得尤其严重。此时，即便是一个高度显著的模型，也很难将观察值正确地进行归类。另一种测量考虑的是如果根据模型对每个观察进行归类，会减少多少归类误差。例如，一个样本容量为 100 的样本，如果有 70 个观察值为众数类别，而模型正确地归类了 80 个观察值，那么误差就减少 33%（即如果按照众数类别进行猜测，我们会犯 30 次错误，而按照模型进行猜测，我们会犯 20 次错误，模型减少了 30 次犯错可能中的 10 次）。

——Timothy M. Hagle

（叶鹏飞译　高勇校）

参考文献

Aldrich, J. H., & Cnudde, C. F. (1975). Probing the bounds of conventional wisdom: A comparison of regression, probit, and discriminant analysis. *American Journal of Political Science*, *19*, 571-608.

Aldrich, J. H., & Nelson, F. D. (1984). *Linear probability, logit and probit models.* Beverly Hills, CA: Sage.

Hagle, T. M., & Mitchell, G. E., II. (1992). Goodness-of-fit measures for probit and logit. *American Journal of Political Science*, *36*, 762-784.

Liao, T. F. (1994). *Interpreting probability models: Logit, probit, and other generalized linear models.* Thousand Oaks, CA: Sage.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics.* Cambridge, UK: Cambridge University Press.

McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, *4*, 103-120.

投射技术（Projective Techniques）

投射技术被定义为一种工具，要求被试者去看一个模糊的刺激并说出它的结构——例如，说出他或她在墨迹上看到了什么。研究者认为被试者会将自己的人格“投射”进这个回答中。

将投射技术与客观测验进行对比是非常有用的。前者接收到模糊的刺激，被试者可以在墨迹上看到从土豚到斑马的任何东西。这种响应的巨大自由提高了回答的投射性质，但很难对其进行量化。在客观测验中，被试者可能必须在关于自我陈述的“是”或“否”两个答案中选择一个。这种限制性的回答提高了统计分析的可能性。因此我们可以说，根据它们的优势和劣势，投

射和客观测验是相互对立的。这将导致一些心理学家认为,它们提供了不同类型的信息,投射测验更多地提供个体的独特性是什么(特殊知识【Idiographic】资料),而客观测验更多地告诉我们在人格特质方面,个体与一个群体如何进行比较(通则知识【Nomothetic】资料)。

投射技术提供的信息类型的一个例子是,病人在看过墨迹上的“四只粉色鹅”之后的联想。“小时候,我在叔叔的农场工作,他把鹅杀死之后,鹅的羽毛会随着血液变成粉红。最近我想起了他折磨过我。有趣的是——他有四个女儿。”客观测验提供信息的例子是,在与常态样本进行对比时,发现在测量抑郁的项目上得分非常高,表明病人当前患有临床抑郁症。

投射技术通常由于在心理测量方面较弱而受到学术心理学家的批评。尽管如此,它们在临床方向的心理学家中仍然很受欢迎。

还有一些更知名的投射技术,其中包括罗夏技术(Rorschach technique),它是由 10 个墨迹组成,被试者必须说出在上面看到了什么。第二个受欢迎的投射技术是主题统觉测验(Thematic Apperception Test,TAT),被试者必须针对人际交往场景的模糊画面建构故事。尽管 TAT 和类似技术都有计分方法,但通常还是凭印象作出分析而不是使用计分。第三个常见的投射技术是 DAP 测验(画人测验,Draw-A-Person),被试者画出一个人以及房屋、树和家庭等这类事物。第四个投射技术是不完整语句测验(incomplete sentences test),被试者必须完成一个句干,比如“大多数女人……”。这种项目的答案通常是凭印象进行分析(参见语句完成测验【Sentence Completion Test】)。

投射技术研究的一些近期趋势,包括不同计分方法的标准化和同时作出的使投射技术更具心理测量学特点的努力(Exner, 2002),为少数人群体设计的替代技术的发展(Costantino, Malgady, & Vazquez, 1981),以及将这些技术运用为心理治疗过程的一个实际部分(Aronow, Reznikoff, & Moreland, 1994)。

——Edward Aronow

(叶鹏飞译　高勇校)

参考文献

Aronow, E., Reznikoff, M., & Moreland, K. (1994). *The Rorschach technique.* Needham Heights, MA: Allyn & Bacon.

Costantino, G., Malgady, R. G., & Vazquez, C. (1981). A comparison of the Murray-TAT and a new Thematic Apperception Test for urban Hispanic children. *Hispanic Journal of Behavior Sciences*, *3*, 291-300.

Exner, J. E. (2002). *The Rorschach: A comprehensive system* (Vol. 1). New York: John Wiley.

校验步骤(Proof Procedure)

校验步骤是谈话分析【Conversation Analysis, CA】的一个主要特征,在校验步骤中,对谈话中某个话轮(turn)表达什么的分析,是基于或是涉及它在下一个说话者话轮中是如何被响应的。因此,一句话被认定是一个“要求”,并不是因为它的内容或语法,

也不是因为分析者直觉上认定如此，而是因为它在下一个话轮中被如此对待。它是在互动中被建构成一个“要求”的，是通过对方的遵从或不遵从或质疑或其他反应，才被建构为“要求”的。

校验步骤不只是一种对已作出的分析的效度进行检查的方法。更确切地说，它是对话分析最初如何开展的一个基本原则。它能够作为对话分析方法论的组成部分，因为它首先是一个参与者的步骤，对于对话参与者自身完成对话并检查他们关于自身言行的理解来说是不可或缺的。参与者基于他们前一个话轮中的言行而开展下一轮的表现，分析员因此有理由使用下一个话轮来开展分析（Sacks, Schegloff, & Jefferson, 1974）。因此，校验步骤是与对话中的话轮转换这一基本现象联系在一起的，包括“条件性关联（conditional relevance）”和“修正”（Schegloff, 1992）；参见对话分析的词条。

这里需要注意的重要一点是，校验步骤不是建立或证明说话者可能“实际上意图”表达什么的方法。下一个话轮可以被理解为显示或暗示参与者对前一个话轮的聆听，但是这种聆听可能会改变或颠覆任何意图，而不是单纯地站在听众立场以最大努力去理解所闻（参看 Schegloff, 1989, p. 200, 关于“下一个行动如何能够重铸先前行为”）。但这并非是下一个话轮校验步骤的失败，因为对话分析的目标不是心理学地理解说话者的意图，而是对话中的互动如何运作的方法（在常人方法论【Ethnomethodology】的意义上）。此外，在已经听到第一个话轮是如何被回应之后，参与者可以选择“修正”在随后的“第三个话轮”中表现出来的理解。于是，我们了解到对话分析的研究目标——不是说话者试图表达的意义，而是他们实现有序谈话的互动过程。意义被看成是互动的完成，而检验步骤直接将我们带入参与者表达意义的互动的偶然性特征，这也是对谈话记录中的细节进行经验研究的基础。

——Derek Edwards

（叶鹏飞译　高勇校）

参考文献

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696-735.

Schegloff, E. A. (1989). Harvey Sacks—lectures 1964-1965: An introduction/memoir. *Human Studies*, *12*, 185-209.

Schegloff, E. A. (1992). Repair after next turn: The last structurally provided defence of intersubjectivity in conversation. *American Journal of Sociology*, *97*(5), 1295-1345.

倾向值（Propensity Scores）

倾向值被定义为在一系列协变量取值给定的条件下，一个个体进入实验组而非对照组的概率【Probability】。倾向值通常被用于因果【Causal】推断的情况，目的是通过比较两组被试者的结果：暴露在干预因素中的一组（实验组）和未暴露的另一组（对照组），来评估某种干预因素（新的实验）的效果。

罗森鲍姆和鲁宾（Rosenbaum & Rubin, 1983）创造了倾向值这一术语，指出了对倾

向值的子分类、成对匹配【Matching】和基于模型调整都会得到实验效果的无偏【Unbiased】估计，只要实验分配是基于倾向值中的协变量【Covariation】的。子分类涉及在倾向值的基础上，将观察值（实验组和对照组）分成数据块。在每一个数据块中，所有协变量分布在实验组和对照组中是一样的，至少在大样本【Sample】中是如此。这个结果可以通过统计检验（如 t 检验【t-Tests】）来进行检查。随机化【Randomized】研究中的因果推断方法可以用在每一个数据块内部。成对匹配法是为每一个实验组成员选取一个倾向值的数值最接近的对照组成员。然后对匹配样本进行分析。在匹配组中，由于外推【Extrapolation】较少，所以建模所需要的假设（如线性）要比在全部实验和控制样本中会得到更好的满足。

在倾向值给定的条件下，实验分配独立于协变量，这意味着在相同的倾向值数值上，观察到的协变量的分布在实验组和对照组中是一样的。这个结果表明，在匹配或子分类中单维倾向值的作用取代了多维的全部协变量。

在随机化试验中，倾向值是已知的。但是在观察研究中，则必须从数据中进行估计。倾向值通常用 Logistic 回归【Logistic Regrssion】来估计。也可使用其他分类方法如分类回归树【Cart】或判别分析【Discriminant Analysis】。德赫贾和沃赫拜（Dehejia & Wahba，1999）提供了一个使用倾向值的观察研究实例，包括估计的细节。

通过运用倾向值中的子分类和/或划分数据块来设计的观察研究【Observational Study】，在两个重要方面都与随机化试验类似。第一，分析都在观察到的协变量分布是相似的实验组和对照组中完成。在随机化试验中，作为随机化的直接后果，所有协变量的分布在实验组和对照组中都是一样的。倾向值则是在观察研究中复制这种情况。

运用倾向值来设计的观察研究类似于随机化试验的另一个方面，是在设计阶段不使用结果变量的数据。倾向值的估计和评估、匹配样本和/或子类的形成，都没有涉及结果。鲁宾（Rubin，2001）更详细地讨论了在观察研究如何使用倾向值。

——Elizabeth A. Stuart

（叶鹏飞译　高勇校）

参考文献

Dehejia, R. H., & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, *94*, 1053-1062.

Rosenbaum, P. R. (2002). *Observational studies*. New York: Springer-Verlag.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41-55.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, *2*, 169-188.

Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual Reviewof Sociology*, *25*, 659-706.

误差减少比例(Proportional Reduction of Error, PRE)

假设我们有两个模型来预测变量 Y,并且对于每一个模型,我们对预测【Prediction】的不精确度都有一些测量。同时,假设模型 1 更为"简单"或信息更少,例如,对 Y 进行预测时不知道任何 Y 的预测因子的信息;模型 2 有更多的信息,运用 Y 的一个或多个预测因子的信息来预测 Y 值。关于模型 1 不精确度的测量即 E_1,可以是一个对于 Y 的观测值与 Y 的预测值之间变差【Variation】的测量,这是一个误差【Error】数量的实际计数,或者,当相同(简单的)规则下(例如,给出每个类别中的个案数,随机猜测哪个个案属于哪种类别)不同的预测会产生 E_1的不同数值时,也可以是基于模型 1 中预测规则的误差期望值。对于模型 2 不精确度的测量即 E_2,是第二个模型中 Y 的估计值的变差,或者是用第二个模型作出预测的误差实际值(不是期望值,因为信息更多的模型 2 应该产生每个个案的唯一预测)。

误差比例的改变是$(E_1-E_2)/E_1$这一比率。正如梅纳德(Menard, 2002)所描述的,由于某些测量的错误,简单的模型 1 预测可能会比信息更多的模型 2 预测更好,在这种情况下,误差比例是增加的。然而,如果$(E_1-E_2)/E_1$的值是正数,表明我们使用模型 2 能够比模型 1 更好地预测 Y 值,运用模型 2 产生的结果是误差减少比例 PRE = $(E_1-E_2)/E_1$。将 PRE 乘以 100 可以得到一个误差减少的百分比,这是运用模型 2 代替模型 1 进行预测能减少的误差百分比。对 PRE 测量的清晰解释让科斯特纳(Costner, 1965)及其他人认为,在社会科学研究中,PRE 测量应优先于其他相关测量方法。

PRE 测量实例包括适用于名义变量之间双变量关联的哥德曼和克鲁斯卡尔的 τ_{YX} 和 λ_{YX},其中 Y 为被预测的变量,X 为预测因子(Reynolds, 1984);适用于定序【Ordinal】变量之间双变量关联【Association】的肯德尔系数 τ_b 的平方$(\tau_b)^2$;以及适用于定比变量的皮尔逊 r 系数的平方 r^2。对于由诸如 Logistic 回归【Logistic Regression】或判别分析【Discriminant Analysis】所产生的预测表,PRE 测量包括 λ_p,τ_p 和 φ_p,类似于列联表【Contingency Table】的相应测量但并不完全相同(Menard, 2002)。对于特殊的 2×2 列联表(不针对更一般的),几个这样的 PRE 测量会得到相同的数值:$r^2=\tau_b^2=\tau_{YX}^2$。在一个因变量有超过 1 个预测因子的模型中,大家所熟悉的最小二乘【Ordinary Least Squares】回归分析的 R^2 以及 Logistic 回归分析的似然比【Likelihood Ratio】R_L^2 也都是 PRE 测量(Menard, 2002)。

——Scott Menard

(叶鹏飞译　高勇校)

参考文献

Costner, H. L. (1965). Criteria for measures of association. *American Sociological Review*, *30*, 341-353.

Menard, S. (2002). *Applied logistic regression analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Reynolds, H. T. (1984). *Analysis of nominal data* (2nd ed.). Beverly Hills, CA: Sage.

Wilson, T. P. (1969). A proportional reduction in error interpretation for Kendall's tau-b. *Social Forces*, *47*, 340-342.

协议(Protocol)

在社会科学方法文献中,协议有几种用法,但是一般而言所有这些用法的核心含义都是"对一项研究或分析工作的程序进行的详尽陈述"。从精神病学到社会学,从政治学到传染病学,所有这些领域中都会遇到这个核心含义(但是在计算机和信息科学文献中,协议也是一个常见的术语,但在这些文献中,它的意义是不同的,是指在不同的处理单元中进行信息交换的一套规则。在此处理单元通常是指计算机,但也可以指有感知能力的人)。

这个术语在社会科学中早期的用法之一是在用问卷进行的研究中(无论是在广泛的抽样调查中或者焦点小组访谈中),在此访谈协议指的就是问卷【Questionnaire】或访谈安排【Interview Schedule】。当然这里对程序的详尽陈述只限于问题的顺序和措辞。

另一种较早的用法是在对解决问题的人工智能【Artificial Intelligence】模型中,其中"出声思维协议"或言语协议【Verbal Protocol】被首先用来比较人的行为和模型行为,之后对于理解人们怎样思维提供了帮助(Ericsson & Simon, 1984; Newell & Simon, 1961, 1971)。出声思维协议,再现了参与者在解题过程中是如何一步一步进行思考的。这些叙述就是调查研究的原始资料。也许因为对言语协议的详尽分析工作量太大,我们在多数社会科学中发现这些协议很少被使用。但是在心理学的某些领域中,它们相当重要(Carroll, 1997)。

目前协议一词在社会科学中越来越多的用法是从医药学中借用来的,它常用来和处理【Treatment】或研究一起使用。这种意义和研究设计【Research Design】很相近,虽然医药学中的处理协议比社会科学中的非实验研究设计遇到的更详细得多。特别是在临床实验中,一个协议涵盖了所有程序的方方面面,以便让不同时点进行的资料收集都能够标准化。与此不同,社会科学中非实验研究设计中标准化的领域多数集中在测量程序上。例如,在许多实地研究中会努力地把提问和答案进行标准化,在文本处理中也会对编码【Coding】进行详尽的标准化。

实验【Experiment】的本质就要求有详尽的研究协议或设计协议,以便调查者能够在尽可能稳定的情况下进行多个实验活动。因此,在社会科学中实验方法的广泛使用也会使得研究协议被广泛地应用。这些协议的发表也会使得精确的复制【Replication】比以往更加常见,这也是对研究工作的重要贡献。

——Douglas Madsen

(高勇译校)

参考文献

Carroll, J. M. (1997). Human-computer interaction: Psychology as a science of design. *Annual Review of Psychology*, *48*, 61-83.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge: MIT Press.

Newell, A., & Simon, H. A. (1961). Computer simulation of human thinking. *Science*, *134*, 2011-2017.

Newell, A., & Simon, H. A. (1971). *Human problem solving.* Englewood Cliffs, NJ: Prentice Hall.

代理回答(Proxy Reporting)

代理回答指的是进行社会调查时遇到的两种彼此相关的情况。如果调查研究者按照某些事先规定的标准,某个被抽中的人不能回答调查问题,那就要用代理人来代替他回答。更一般而言,让一个人讲述其他人的信息,这种情况都可以被定义为代理回答。

——Tim Futing Liao
(高勇译校)

代理变量(Proxy Variable)

代理变量就是用来测量一个无法直接观察的量的变量。虽然代理变量不是对这个数量的直接测量,但是一个好的代理变量和不可观察的变量是有密切关联的。在社会科学中代理变量非常重要,经常被使用,因为对某些数量进行直接测量非常困难或根本不可能。

用一个例子可以说明代理变量的重要性以及存在的问题。假如一个研究者想考察智力对收入的影响,同时控制年龄的影响。如果可以知道人的智力,而且假定两者的关系是线性的,那么这种关系就可以用如下多元回归【Multiple Regression】来进行估计:

$$收入_i=\alpha+\beta_1\times智力_i+\beta_2\times年龄_i+\varepsilon_i$$

不同于年龄,智力不是一个可以直接观察的量。因此对上述关系感兴趣的研究者必须用和智力相关的其他测量,例如教育年数或标准化的测试结果。用教育年数作为智力的代理变量:$智力_i=教育年数_i+\delta_i$。这样,研究者就要估计下面的式子:

$$收入_i=\alpha+\beta_1\times教育年数_i+\beta_2\times年龄_i+\varepsilon_i$$

严格地说,β_1 测量的并不是智力和收入之间的协变关系,除非对于所有的 i 来说,$\delta_i=0$。但是因为不可能知道代理变理在多大程度上测量了那个不可观察的量(即 δ_i 的取值是多少),并且因为收入和教育年数的关系能帮助我们理解收入和智力的关系,因此 β_1 通常在解释中被等同于后者。有人建议如果有多种测量办法(如上面的例子所示),就应该去比较不同的设定方法来选择最佳的代理变量(Leamer, 1978)。

统计上,决定是否要用代理变量取决于对两个可能的后果进行的权衡。如果要用代理变量,研究者就必然会遇到与"变量中的误差"相关的传统问题(即所有的参数估计【Parameter Estimation】是没有一致性的)。但是如果不用代理变量,而且无法测量的量又和设定相关,那么没有使用代理变量将会导致遗漏变量偏倚(参见遗漏变量【Omitted Variable】和偏倚【Bias】)。对这种权衡的计量经济学分析表明,一般来说,包括代理变量更为可取,甚至代理变量相对较弱时也是如此。

——Joshua D. Clinton
(高勇译校)

参考文献

Fuller, W. A. (1987). *Measurement error models.*

New York: John Wiley.

Kinal, T., & Lahiri, K. (1983). Specification error analysis with stochastic regressors. *Econometrica*, *54*, 1209-1220.

Leamer, E. (1978). *Specification searches: Ad hoc inferences with nonexperimental data*. New York: John Wiley.

Maddala, G. S. (1992). *Introduction to econometrics*. Englewood Cliffs, NJ: Prentice Hall.

伪 *R* 方(Pseudo-*R*-Squared)

回归【Regression】分析中常用的拟合优度量度【Goodness-of-fit Measure】就是(多元)决定系数【Coefficient of (Multiple) Determination】,也被称为 *R* 方【*R*-Squared】或者 R^2。R^2 本质上测量的是因变量【Dependent Variable】的方差中能够由多元回归【Multiple Regression】模型解释的比例。虽然对于 R^2 的效用仍然存在争议,但是它确实是对于模型表现的一种常用测量。

回归分析者有一系列诊断工具和拟合优度测量。这些工具和测量对于 probit 模型和 logit 模型则不再适用。特别是,不存在可以类比于 R^2 的拟合优度测量。有几种替代测量被称为伪 R^2,在许多计算机程序结果中也会提供。probit 和 logit 分析者面临的一个困难是在多个伪 R^2 中如何选择。在选择伪 R^2 时,首要问题是要用这个伪 R^2 测量来干什么(如对方差【Variance】的解释,假设【Hypothesis】检验,对因变量进行归类等)。

最常见的用法是测量模型能够解释多少因变量的变异。一种常用的方法是由麦凯尔维和扎沃伊纳(McKelvey & Zavoina, 1975)提出的,是用 probit 或 logit 系数来计算潜变量【Latent Variable】的预测值的方差,用 $\mathrm{var}(\hat{y}_i)$ 表示。扰动项的方差为 1,因此总方差就是解释方差加 1。这样,伪 R^2 就可用下式来计算:

$$R^2 = \frac{\mathrm{var}(\hat{y}_i)}{1+\mathrm{var}(\hat{y}_i)}$$

另一种常见的伪 R^2 是由奥尔德里奇和纳尔逊(Aldrich & Nelson, 1984)提出的,使用的是卡方统计量 $-2LLR$,即 $-2\ \ln(L_0/L_1)$,自由度【Degrees of Freedom】为 k, 即用来估计的自变量数量。这个测量需要对数据进行两次拟合,第一次只用常数来估计得到 L_0,第二次用全模型来估计得到 L_1。伪 R^2 用下式计算为:

$$R^2 = \frac{-2LLR}{N-2LLR}$$

其中 N 为观察数。

哈格尔和米切尔(Hagle & Mitchell, 1992)使用模拟【Simulation】数据进行了检验,证明这两种测量都是一般最小二乘【Ordinary Least Squares】R^2 的良好估计。奥尔德里奇和纳尔逊(Aldrich & Nelson, 1984)提出的测量方法误差略小一些,如果因变量的正态分布【Nornal Distribution】假设不成立时,这种测量也更为稳健【Robust】。奥尔德里奇和纳尔逊提出的测量在理论上的局限性在于,它与处于众数类别的观察比例相关。

probit 和 logit 中伪 R^2 的用处是基于一个假定,即存在一个不可观察的连续变量,

但它只能表现为一个可观察的二分变量。例如,不可观察的变量可以是某个投票者选择候选人 A 而非候选人 B 的概率,而可观察的变量是它在 A 和 B 之间的真实选择。不可观察的变量不同,但相对应的可观察变量可能是相同的,因此分析者更为强调模型对观察进行正确分类的能力。使用正确分类率来作为拟合优度测量的问题在于,如果处于众数类别的观察比例较高时,probit 和 logit 就很难将观察正确地归类到非众数类别中。总体而言,综合使用多个测量能够更好地理解模型的表现。

——Timothy M. Hagle

(高勇译校)

参考文献

Aldrich, J. H., & Nelson, F. D. (1984). *Linear probability, logit and probit models.* Thousand Oaks, CA: Sage.

Hagle, T. M., & Mitchell, G. E., II. (1992). Goodness-of-fit measures for probit and logit. *American Journal of Political Science*, *36*, 762-784.

McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, *4*, 103-120.

心理分析方法(Psychoanalytic Methods)

心理分析既是一种临床心理治疗的实践活动,也是一种关于人们精神世界的一套理论体系,包括精神世界的发展及其在各种活动和关系中的表现。经验发现是心理分析的基础,但是这些发现并不认为自己具有客观性,而被认为是病人和分析者的主观产物。有人反对精神分析的资料具有科学性,其理由是精神分析的方法是在咨询室而不是在实验里实现的,得到的是孤立的个案而不是具有普遍性的结果。但是,随着社会科学研究中对于实证方法的挑战,语言学的、定性的、解释学的转向为心理分析在定性方法中开辟了立足之地。

社会科学家通常假设他们的研究对象是理性的、统一的、有明确意向的主体,但心理分析理论有力地质疑了这种明智的主体。这些访谈研究、调查研究、态度量表研究的基本理念认为,被访者是充分了解自己的、无话不谈的,但是心理分析理论认为存在着无意识的冲突和防御,这些对前面的理念提出了怀疑(Hollway & Jefferson, 2000)。这些无意识的内容必定需要人的解释,心理分析概念可以帮助进行这样的解释(例如,如果回忆一件事很痛苦,人们就会否认这件事情)。这类概念已通过临床的心理分析实践得以建立和完善(Dreher, 2000)。

心理分析方法如果要扩展到咨询室以外就需要进行一些修正,在咨询室中研究发现是服从于治疗的目的。临床应用和研究应用都会使用心理分析的解释。在咨询室中,分析者的解释是他们临床方法的一部分,直接提供给病人,病人的情绪反应本身就可检验这个解释是否正确。在用心理分析方法得到解释和研究资料时,需要寻求其他方法来验证这些解释是否有效(例如,在全部文本中寻求正面的例子和反面的例子,或者察觉到可以识别的防御模式)。

自由联想方法是心理分析的一个基础,

它鼓励病人毫无保留地说出自己的想法,其假定是没有经过审查的思想才能引出重要的东西。使用自由联想有助于得到和分析访谈资料(参见自由联想访谈法【Free Association Interviewing】),可以接触到社会生活的某些方面,而那些认为研究对象是了解自己的、无话不谈的方法没有意识到这些侧面(例如,影响家庭和工作关系的一些欲望和焦虑)。移情的概念是指愿望无意识的从一个早期的人(如父母)身上转移到了分析者身上,他们当时在这个人身上体验到这种情感。虽然在非临床的情况下会有所淡化,对于自己和参与者的移情和反移情敏感的研究者可以使用这些作为研究资料(如 Walkerdine, Lucey, & Melody, 2001)。从认识论上来讲,这些概念对于客观性提出了批评,代之以研究者对反思性的理解,这是一个在定量社会科学中很重要但没有被理论化的概念。

访谈和观察都会包括心理分析的影响。克沃勒(Kvale,1999)认为精神分析访谈中得到的知识,对于研究来说是重要的。有两种观察研究使用了心理分析方法,具体而言是民族志的实地研究(Hunt, 1989)和婴儿观察(Miller, Rustin, Rustin, & Shuttleworth, 1989)。在德国,心理分析也被用来进行文化研究。心理分析式的观察最早可以追溯到1940年代的英国,在塔维斯托克研究所对婴儿每周进行一次详细的观察是他们训练的一部分,现在心理分析观察已扩展到更广泛的研究领域中。

——Wendy Hollway

(高勇译校)

参考文献

Dreher, A. U. (2000). *Foundations for conceptual research in psychoanalysis* (Psychoanalytic Monograph 4). London: Karnac.

Hollway, W., & Jefferson, T. (2000). *Doing qualitative research differently: Free association, narrative and the interview method.* London: Sage.

Hunt, J. C. (1989). *Psychoanalytic aspects of fieldwork* (University Paper Series on Qualitative Research Methods 18). Newbury Park, CA: Sage.

Kvale, S. (1999). The psychoanalytic interview as qualitative research. *Qualitative Inquiry*, *5*(1), 87-113.

Miller, L., Rustin, M., Rustin, M., & Shuttleworth, J. (1989). *Closely observed infants.* London: Duckworth.

Walkerdine, V., Lucey, H., & Melody, J. (2001). *Growing up girl: Psychosocial explorations of gender and class.* Basingstoke, UK: Palgrave.

心理测量学(Psychometrics)

心理测量学是对于人与人之间的个体差异进行的测量。测量使用的工具可以是心理测验、问卷、人格量表、观察清单。测量给出的值,既可以是一种心理特质的水平,也可以是对这种水平的排序。这些心理特质包括能力(如语言智力、分析推理、空间感)、技能(如操控指挥交通用的仪表盘,在工作情境中主持会议)、人格特点(如内向、友善、神经质)、态度(如对堕胎、政治右翼、北约的武装干涉的态度)。“约翰比玛丽更内向”这是对一个人格特点进行排序。“卡罗尔的智商足够高,能够参加这个课程”,这

是在比较卡罗尔的认知水平和参加这个课所需要的绝对水平。

通过测验、问卷或其他工具得到的测量值必须是有信度和有效度的。一个测量是有信度的，是说同一种测量程序进行重复测试能够具有稳定性。更广义而言，在不同的测量单元中（一个测验的不同题项），在不同的测量场合中（时点），对于评估学生和客户的不同心理学家，这种稳定性也应当存在。一个测量是有构念效度【Construct Validity】的，是说它对于要测量的特质进行了充分的操作化。这意味着，一个测量工具测量的是人的内向性，它就不应该测量回答者的理想自我。预测效度指的是测量能够对人的未来行为进行准确预测的程度（例如，分析推理能力能够预测一个人是否适于成为一个计算机程序员）。

心理测量学研究和提出了许多数学和统计测量工具。比较老的心理测量工具包括斯皮尔曼提出的经典测验理论；评估量表数据的李克特量表【Likert Scaling】；斯皮尔曼、瑟斯通、汤姆森在1920年代和1930年代提出的因子分析【Factor Analysis】；瑟斯通提出的分析成对比较数据的方法；哥特曼提出的分析有导向数据（dominance data）的量图分析（scalogram analysis）；库姆斯提出的分析偏好数据的展开方法（unfolding method）。概化理论【Generalizability Theory】主要应归功于克朗巴赫，这一方法使用方差分析方法来对于检验结果的来源进行分解。项目反应理论【Item Response Theory】I模型是由洛德、伯恩鲍姆、拉施、莫肯等人在1950年代和1960年代最早提出的一组现代心理测量模型。自此以后，许多研究者接受了这一理论模型，在随后几十年中主导了心理测量学。

从传统上讲，多元分析【Multivariate Analysis】方法一直是心理测量学的一部分。具体而言，因子分析和聚类分析【Cluster Analysis】用来探索构念效度，回归方法用来建立测量题组来进行预测。生理心理测量法，如确定视觉阈限和听觉阈限，主要目标是建立行为的一般法则而不是测量个体差异。

心理测量学还包括建立各种测量工具的技术，如标准的纸笔测试，再如建立题库进行计算机化的测量程序，以便进行更有效的测量。其他应用还包括确定使某些少数群体处于不利的题项（题项运作差异分析），确定在题项得分上具有异常模式的回答人（人—拟合分析）。应用还包括对于回答人在测验中用来解题的认知过程和解题策略的研究（认知建模）。

——Klass Sijtsma

（高勇译校）

* 也可参见项目反应理论【Item Response Theory】。

心理生理测量（Psychophysiological Measures）

心理生理测量指的是心理构念（如心理状态或心理过程）的生理测量方法。研究者在20世纪早期开始用科学方法探索身和心的关系。但是直到20世纪中期生理心理学才从生理学倾向的心理学中独立出来，成为独立的分支学科，在此之前，生理学倾向的心理学家探索的是生理变化（特别是精神生理变化）对行为和精神的影响，他们通常用动物来做实验。与此不同心理生理学家探索的是心理变化对生理结果的影响，他们通

常是用人来做实验。现在心理生理学和生理学倾向的心理学之间的区别已经模糊。两者都被归到更广泛的神经科学中去了。

然而,生理心理学家已对心理学和行为科学作出了影响深远的贡献,对心理状态提供了很多重要和有用的生理测量方法。对心理状态进行生理测量基于一个假定,即对生物过程或生理过程(特别是那些无法用意识来控制的过程)进行观察和记录在方法上更可靠。

生理测量和心理生理测量通常是即时的、隐秘的、连续的。即时的生理测量可以在时间上非常精确。它们可以马上和测量的预期要发生的心理状态或过程联系起来,从而能够马上提供证据,证明这种心理状态或过程的强度或运作,而其他一些事后的测量则存在缺陷,因为记忆是有偏差的。例如,对于快乐进行即时的生理测量,就能避免由于回忆而导致的测量误差。

外在的观察者和被评估的参与者都不能很容易地看到隐秘的生理测量的结果,这样就增强了其真实性。由于生理测量是隐秘的,参与者得不到有意识的反馈,因此参与者也就没有办法控制他们的反应。由于它们具有隐秘性,生理测量就避免了参与者用各种策略来进行印象管理,而这在实验情境中却是可能发生的。例如,对快乐进行隐秘的测量就避免了参与者表现得比他们实际的更快乐或更不快乐。

连续的生理测量能够准确地描绘出起伏变化。实验者可以以非常高的频率来不断进行生理测量。事实上在计算机技术的帮助下生理测量的记录可以达到接近完美的保真度。因此研究者可以记录下生理的起伏变化,这样增加了信息却并没有增加成本。例如,记录下快乐的起伏变化,这样研究者能够考察快乐在一段时期内的趋势变化。即使测量时的集中趋势【Central Tendency】是相同的(如均值、中位值、众值),起伏变化的差异能够给出重要的心理信息。图1中参与者的反应有着相同的中心趋势,但起伏变化各不相同。

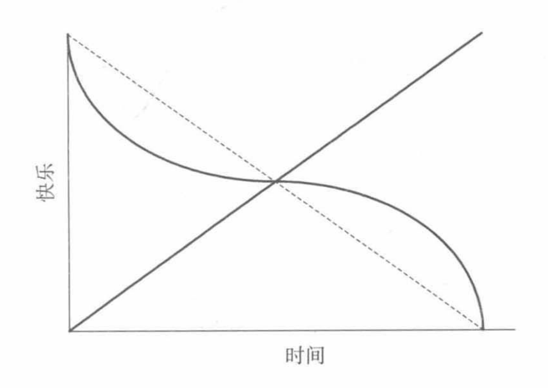


图1 心理生理测量

但是生理测量给出的资料是即时的、隐秘的、连续的,资料具有即时性、真实性、能记录下起伏变化,上述事实并不意味着它们具有构念效度【Construct Validity】。因此,心理生理测量在方法上仍存争议。事实上,上面提到的关于快乐的例子并没有具体说明所用到的生理指标或测度。即使对于快乐研究进行文献梳理也会得到很多用到的生理测量,例如肌电活动、心率、大脑皮层的血流量等,把这些生理测量和快乐联系在一起的理论并不是那么清楚。

和心理研究中所用到的测量相似,生理测量也必须具有构念效度。其他学者也已讲过(Blascovich, 2000; Cacioppo, Tassinary, & Berntson, 2000),心理构念和其指标之间的关联应具有恒定性(invariance),无论这个指标是心理指标还是生理指标。也就是说这种构念和指标应一起发生,一个东西出现另一个东西也要随之出现,反之亦然。但是心理构念和生理指标之间的恒定性是很难达到的,因为有些心理构念(如快乐、感情、冒险、偏见、好学)是很难界定清楚的,而

身体是一个服务于一般目的的系统而不是服务于单个目的的系统。

因此和非生理测量一样,研究者需要一种测量理论来建立恒定性假设。对于生理测量来说,测量理论必须基于心理生理理论,把这种构念及其指标联系起来。这绝非易事,但是已经有一些有力的理论为许多心理构念提供了必要的支持,包括如感情、动机和认知等心理领域中的一些构念(相关评论,请参见 Blascovich, 2000; Cacioppo, Tassinary, & Berntson, 2000)。

例如,卡乔波等人(Cacioppo, et al., 1986)建议用肌电图(electromyographic, EMG)技术来捕捉颧大肌和皱眉肌的反应来辨别正面感情和负面感情。其心理生理基础涉及达尔文的理论,认为用于社会互动的脸部表情在进化过程中非常重要。他们推理说,正面感情会伴随着颧大肌(微笑肌)肌电的增加,也会伴随着皱眉肌肌电的减少;负面感情则有相反的模式。随后他们用实验证明了这种模式的存在。

布拉斯科维及其同事(如 Blascovich & Tomaka, 1996)建议用心电图(electrocardiographic, ECG)、心电阻抗(impedance cardiographic, ZKG)、血压监控技术来捕捉心血管反应模式,用于表明人遇到挑战和威胁的动机和状态。这种做法的心理生理基础涉及丁斯特比尔(Dienstbier, 1989)提出的生理僵硬的精神内分泌理论。布拉斯科维及其同事将挑战界定为一种良性动机状态(即资源足以应付情境要求),挑战会伴随着交感肾上腺髓质(Sympathetic Adrenal Medullary, SAM)促发的心肌收缩和心输出量(cardiac output)的增加,与此同时血管舒张;而威胁是一种恶性的动机状态(即资源不足以应付情境要求),威胁会伴随着垂体—肾上腺皮质(Pituitary Adrenal Cortical, PAC)对于 SAM 的抑制,结果导致心肌收缩增加,但是心输出量和血管张力(vasomotor tone)不会发生变化。

许多基于理论的对心理构念的生理指标不断出现,这类测量越来越多。随着诸如脑成像这样的新技术不断出现,人们越来越认识到生理测量对于心理研究的用处,我们期待会出现越来越多、越来越好的生理测量指标。

——Jim Blascovich

(高勇译校)

参考文献

Blascovich, J. (2000). Psychophysiological indexes of psychological processes. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 117-137). Cambridge, UK: Cambridge University Press.

Blascovich, J., & Tomaka, J. (1996). The biopsychosocial model of arousal regulation. In M. Zanna (Ed.), *Advances in experimental social psychology* (pp. 1-51). New York: Academic Press.

Cacioppo, J. T., Petty, R. E., Losch, M. E., & Kim, H. S. (1986). Electromyographic specificity during simple physical and attitudinal tasks: Location and topographical features of integrated EMG responses. *Biological Psychology*, *18*, 85-121.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2000). *Handbook of psychophysiology* (2nd ed.). Cambridge, UK: Cambridge University Press.

Dienstbier, R. A. (1989). Arousal and physiological toughness: Implications for mental and physical health. *Psychological Review*, *96*, 84-100.

民意研究(Public Opinion Research)

民意研究涉及对人们的想法和行为的概念化【Conceptualization】与测量【Measurement】。自从这个领域伴随调查研究共同发展起来之后,从业者对于民意的意义和测量就有着不同看法,对于"民意研究"的研究至今仍然是一个活跃和充满争议的领域。当然,也可以认为"民意"除了回答调查问题以外,也可通过诸如罢工和示威游行等集体运动表达出来,但调查【Survey】和测验仍然是测量和报告民意最重要的方式。

在概念上,最关键的一点是,民意是一个个体概念还是一个集体【Aggregation】概念。虽然个人会有自己的态度和观点并且表达出来,但民意是不是就是这些个人观点的汇总呢?如果每个人都有自己的观点,但是这些观点没有被测量和汇总,他们也没有和别人交流这些观点,那么事实上还会基于这些观点发生集体行动吗?另一个关键之处在于,如果民意没有被表达,它是否还存在。这个道理被用来区分"测量出"的民意和"考虑过"的民意。有这样的可能,不管问题的质量如何,提问的群体是谁,个人对这个问题进行了多少考虑,用一个调查问卷测量出来的东西,就被人称为"民意"。测量出的民意反映的是个人面对问卷作出的回答,无论他们对这个问题是否关注,而考虑过的民意是经过长时间深思熟虑的最终结果。

通常民意都是调查的结果,要从一个相关总体【Population】中抽取代表性样本【Representative Sample】进行测量。通常它要反映的是社会中或某个选区中所有成人的态度。有些研究者认为,对于民意还有其他测量方法,如投票、示威或者给报纸写信。

民意研究关注两个重点领域:民意的概念维度和调查方法对意见表达的影响。从民意的角度看,大多数人群是分层的,有一小部分政治精英,对于问题有复杂和精心组织的观点。他们对调查问题的回答,在意识形态上是一致而周详的(Converse, 1964)。研究表明,这些精英除了通过媒体表达他们的话语外,还潜移默化地影响大众对于具体问题重要性的看法以及应该在什么领域来考虑这些问题。一般市民在评估这些事情时,精英的政党倾向和政治意识形态的信息很有用,有助于他们不用费力气学习和吸收新的信息就能进行判断(Lupia, McCubbins, & Popkin, 1998; Popkin, 1991)。但是信息不是民意的唯一基础;其他社会心理因素,如刻板印象也会影响民意的形成。对于种族和性别的态度尤其如此,对于依靠福利生活的穷人的看法也是如此。

民意的另一个方面是态度的强度,在政治领域,少数人的强烈感受比所有人的意见分布更能解释政策和立法的变化。如果这种态度强度和某些政治行为有关,如抗议、投票、写信,这一点就尤其重要。

从方法的角度看,民意研究者关心的一点是,被访者经常出于礼貌或者社会惯例来回答问题。结果他们就会随便作答。使用封闭式问题【Closed-Ended Question】,特别是如果不提供"不知道"或"不清楚"这样的选项时,最容易得到这样的一些

回答。

在选择回答对象时可能会出现偏差【Bias】,越来越多的方法问题与此有关。这些问题包括在网页上回答问题的被访者是自我选择的,标准的随机拨号抽样【Random Digit Dialing, RDD】中没有包括普遍使用的手机号码,回答率越来越低等。

由于人们关心民意的意义和调查方法如何影响民意测量,民意研究在未来将是在实质上和方法上都得到关注的一个领域。

——Michael W. Traugott

(高勇译校)

参考文献

Converse, P. E. (1964). The nature of belief systems in mass publics. In D. Apter (Ed.), *Ideology and discontent* (pp. 206-261). New York: Free Press.

Lupia, A., McCubbins, M. D., & Popkin, S. L. (Eds.). (1998). *Elements of reason*. Cambridge, UK: Cambridge University Press.

Popkin, S. L. (1991). *The reasoning voter*. Chicago: University of Chicago Press.

Price, V. (1992). *Public opinion*. Newbury Park, CA: Sage.

Zaller, J. R. (1992). *The nature and origins of mass opinion*. Cambridge, UK: Cambridge University Press.

目的抽样(Purposive Sampling)

定性研究中的目的抽样是指根据分析和理论需要,有意选取有某种特征的研究对象。因为在研究开始时,研究者并没有充分掌握这个现象,样本的性质并不总是能够提前确定。通常是在资料分析的过程中,研究者会意识到它需要去访问当初认为与此无关的被访者。例如,马丁在研究婴儿猝死综合征时发现,有必要研究首先应对问题的人群之间(急诊科大夫和急救人员)的责任划分问题。他需要去访问这些专业人员,而不只是像当初想的那样,只访问孩子的父母就够了。

目的抽样的类型包括提名抽样、滚雪球抽样【Snowball Sampling】(已经参加研究的人会帮助找人来参加研究)和理论抽样【Theoretical Sampling】(随着研究进展根据分析获得的信息来有意寻找合适的参与者)。

在提名抽样或滚雪球抽样中,研究者会找到一个"合适"的参与者,在访谈结束之时请求参与者推荐另一个可能参与研究的人。这样抽样是沿着一个社会网络进行的。如果研究的人群很难找,或者不会自愿参与研究,或者对招聘参与者的广告有反应,这时提名抽样或滚雪球抽样就特别有用。参与者可能因为有非法行为(如吸毒)而感到羞愧或害怕惩罚,参与者可能是属于一个封闭的群体(如飞车帮),参与者可能有隐私行为或者有某种病而可能会被污名化,这时用其他方式很难找到这种被访者,滚雪球抽样就是非常有用的策略。取得第一个参与者的信任,让他向其他群体成员保证这个研究没问题,就能接触到用其他方式接触不到的被访者。

使用理论抽样的研究者也会用到提名抽样。他们会用一个提名抽样表来请人推荐有特定经历或特定知识的参与者,以便推进他们的分析。

——Janice M. Morse

(高勇译校)

参考文献

Martin, K. (1998). *When a baby dies of SIDS: The parents' grief and searching for a reason*. Edmonton, Canada: Qual Institute Press.

皮格马利翁效应(Pygmalion Effect)

皮格马利翁效应有广义和狭义之分。广义而言是指人际期待效应或者人际自我实现预言。这些术语指的是研究发现,一个人期待其他人有什么样的行为,事实上其他人就会有什么样的行为。狭义的皮格马利翁效应是指教师对于学生表现的期待效应。

对于皮格马利翁效应最早的实验【Experiment】研究是用人来作为研究参与者的。实验者要求参与实验的人对于一些肖像照片进行打分,其中一半的实验者被引导期待得到很高的照片得分,另一半被引导期待得到较低的照片得分。在许多研究中,期待得到较高的照片得分的实验者,结果就得到了较高的照片得分;期待得到较低照片得分的实验者,结果就得到了较低的照片得分(Rosenthal, 1976)。

研究者为了在实验室中调查这种人际期待效应的普遍性,用动物做了两个实验(Rosenthal, 1976)。有一半的实验者被告知,他们的小白鼠特别擅长走迷宫(斯金纳盒子);另一半被告知,他们的小白鼠不擅长走迷宫。在两个实验中,如果研究者被引导认为他们的小白鼠更善于学习,结果他们的小白鼠学习能力就更强。

“皮格马利翁实验”是上述研究的直接产物。如果给予小白鼠较高期待,它们都能变聪明,那么理所当然,老师给予孩子更高的期待时,他们也能变得更聪明。在一个研究中(Rosenthal & Jacobson, 1968, 1992),所有的孩子都进行了非语言能力的智力测试,伪称这个测试能够预测智力“发展”。学校一共 18 个班,6 个年级,每个年级 3 个班。这 3 个班分别由能力强的、一般的、能力弱的孩子组成。从每个班随机选取了 20%的孩子组成了实验组。然后告诉每位老师他们班上进入实验组的孩子的姓名,并且说这些孩子在智力发展测试中得分很高,他们会在未来八个月中智力上有很大的进步。但实际上,实验组和参照组孩子们之间的唯一区别只存在于老师的头脑中。

八个月后,学年结束时,对所有的孩子进行了相同的智力测试。整体而言,那些被老师认为会有很大进步的孩子,比其他孩子得到了更大的进步。

共有约 500 个实验,对于人际期待效应现象或广义的皮格马利翁效应进行了研究,通常得到的效应强度是很明显的:如果我们被引导期待别人会有某种行为,65%的情况下我们会得到这种结果,如果我们不被引导期待这种行为,则只有 35%会得到相应的结果。

——Robert Rosenthal

(高勇译校)

* 也可参见实验期待效应【Experimenter Expectancy Effect】、调查者效应【Investigator Effects】。

参考文献

Blanck, P. D. (Ed.). (1993). *Interpersonal expectations: Theory, research, and applications.* New York: Cambridge University Press.

Eden, D. (1990). *Pygmalion in management: Productivity as a self-fulfilling prophecy.* Lexington, MA: D. C. Heath.

Rosenthal, R. (1976). *Experimenter effects in behavioral research: Enlarged edition.* New York: Irvington.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom.* New York: Holt, Rinehart & Winston.

Rosenthal, R., & Jacobson, L. (1992). *Pygmalion in the classroom: Expanded edition.* New York: Irvington.